UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

Big Data : Est-il possible de se cacher du monde numérique ? Par quels moyens et avec quels risques ?

Auteur : Quentin Sellier
Promoteur : François Fouss
Année académique 2018-2019
Master [120] Ingénieur de gestion, à finalité spécialisée

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | www.uclouvain.be/lsm

Résumé

Dans ce travail, nous considérons tout d'abord que « se cacher du monde numérique » équivaut à ne pas laisser de traces numériques que l'on puisse lier à notre identité. Dans la revue de la littérature, l'apport de Louise Merzeau, Fabrice Rochelandet et Grazia Cecere nous pousse à poser l'hypothèse qu'il n'est pas possible de se cacher du monde numérique.

Pour vérifier cette dernière, nous dressons un inventaire des traces que nous laissons, en passant par les smartphones, les ordinateurs, certains services web, les réseaux sociaux et des éléments sortant du cadre du web. Par après, cet inventaire est repris pour essayer de trouver des solutions permettant de limiter nos traces, via des outils, habitudes et techniques. Dans cette partie, nous voyons déjà que certains éléments sont particulièrement compliqués à solutionner, comme par exemple le fournisseur d'accès internet, qui capte des données même si nous utilisons de puissants outils pour nous cacher, ou le smartphone, un indispensable de notre société laissant de nombreuses traces.

Nous mettons ensuite à profit ces outils dans une expérience ayant comme objet central la minimisation de l'emprunte numérique d'un individu durant une certaine période. A la fin de celle-ci, nous faisons valoir notre droit d'accès à nos données personnelles auprès de plusieurs entreprises pour que ces dernières envoient toutes les informations qu'elles possèdent à notre sujet. Grace à cela, une vérification scientifique est réalisée en observant si certaines entités ont capté des traces durant la période de cette expérience. C'est le cas pour plusieurs d'entre elles, analyse permettant de valider l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique. Après cela, ces résultats sont validés lors de l'interview d'un expert du milieu et une enquête nous aide à objectiver certains aspects de ce travail.

Finalement, et en partie grâce à cette expérience, nous passons à une partie concernant les limites, difficultés et risques du comportement d'une personne souhaitant minimiser ses traces. Nous y voyons qu'une telle expérimentation n'aurait pas pu être poursuivie indéfiniment et que des données supplémentaires auraient inévitablement été générées, ce qui confirme la validation de l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique. Il ne faut également pas oublier les risques pour les entreprises, qui seraient presque toutes impactées par de tels comportements, principalement à cause de la fin de la publicité ciblée, des systèmes de recommandation et une faible productivité de leurs employés.

J'adresse mes remerciements aux personnes qui m'ont aidé dans la réalisation de ce mémoire.

En premier lieu, je tiens à remercier mon promoteur, le professeur François Fouss, pour son suivi et ses conseils avisés durant la production de ce travail.

Ensuite, je remercie particulièrement mes proches pour leur aide durant toutes les étapes de cet exercice, notamment pendant l’expérience réalisée et lors de la relecture.

Table des matières

Liste des illustrations

Liste des annexes

Introduction

« 90% des données dans le monde ont été créées durant les deux dernières années »[1]. Il s'agit d'une statistique publiée en 2016 par le géant de l'informatique IBM. Cependant, pourquoi voit-on cette génération exponentielle des données ? Encore selon IBM, les entreprises mettant l'expérience des utilisateurs au centre de leur business ont pu profiter d'une croissance de leur chiffre d'affaire supérieure de 14 points de pourcentage par rapport aux entreprises délaissant cet aspect. En effet, les utilisateurs attendent de plus en plus des entreprises. Ils veulent que celles-ci connaissent leurs préférences, leurs habitudes et leurs besoins afin que leur expérience soit la plus agréable possible. C'est ainsi devenu une réelle nécessité pour ces acteurs.

Pour réaliser cela, les plus grands groupes peuvent profiter d'une technologie évoluant rapidement et permettant la récolte de plus en plus d'informations, elles-mêmes toujours plus complexes. Historiquement, seules quelques données telles que le nom, le numéro de téléphone et certaines transactions pouvaient être relevées et utilisées. Maintenant, différents éléments comme les smartphones, les réseaux sociaux et le commerce en ligne ont permis d'évoluer dans la manière de récolter et traiter nos informations[2]. A titre d'illustration, nous créions chaque jour en 2016 2,5 quintillions d'octets de données[3], soit $2,5 * 10^{12}$ Go. Celles-ci peuvent être simples, telles que des noms, adresses mails, adresses postales, etc., mais, surtout, des informations complexes de transactions, comportements, réactions à des stimuli, etc. permettant une utilisation bien plus poussée. Avec une telle importance accordée à la récolte et au traitement des données personnelles, il est ainsi légitime de se demander s'il est toujours possible de se cacher du monde numérique.

[1] IBM (auteur anonyme), 2018, « 10 Key Marketing Trends for 2017 », consulté le 23/07/2018, En ligne https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

[2] Martinez I., 2018, « Finding Hidden Customer Behavior Patterns Using Big Data Analytics », consulté le 25/03/2018, Dzone, En ligne https://dzone.com/articles/find-hidden-customer-behavior-patterns-using-big-d

[3] IBM (auteur anonyme), 2018, « 10 Key Marketing Trends for 2017 », consulté le 23/07/2018, En ligne https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

Tout d'abord, il semble important de préciser cette question de recherche. Nous parlons en effet du « monde numérique » et du « *Big Data* ». Le monde numérique est un concept assez large qui regroupe la totalité des éléments passant par un processus numérique, soit par exemple les milliards d'objets connectés ou même les appels téléphoniques[4]. La définition de trace numérique, soit une trace passant par un processus numérique, est donc également très large et il est impossible de ne pas en laisser. Ainsi, quand nous parlerons de « trace numérique » dans la suite de ce mémoire, nous ferons plutôt référence à une trace qui est récupérée et liée à un profil. Par exemple, payer en liquide un ticket de parking sans indiquer l'immatriculation de notre véhicule revient à laisser une trace numérique mais celle-ci ne peut pas être liée à notre identité.

Le terme *Big Data*, quant à lui, a longtemps été définit par ses 3V malgré une signification pourtant très large tant celle-ci caractérise les informations, la technologie, les méthodes et son impact. Ainsi, une définition plus complète est celle de De Mauro et al. (2015) selon qui « Le *Big Data* représente les actifs d'informations caractérisés par un Volume, une Vélocité et une Variété tellement grande qu'ils nécessitent une technologie et méthode analytique spécifiques pour leur transformation en valeur ».

Enfin, « se cacher » peut avoir plusieurs significations. Dans ce travail, et étant donné les deux termes expliqués précédemment, nous considérons que « se cacher du monde numérique » fait référence au fait de ne laisser aucune trace numérique qui soit liée à notre identité.

Ensuite, étant donné la nature d'un tel sujet, il est important de cadrer ce mémoire. Ce travail étant réalisé dans le cadre d'études d'ingénieur de gestion, nous nous concentrons plutôt sur le monde des entreprises. Cela comprend plusieurs remarques majeures à relever : nous ne nous pencherons pas sur l'implication d'un état, que ce soit pour une utilisation administrative ou une éventuelle surveillance, et nous ne poserons pas la question des raisons qui poussent à individu à se cacher.

[4] Silog (auteur anonyme), 2015, « Le monde numérique, c'est quoi ? », consulté le 11/01/2019, En ligne https://www.silog.fr/le-monde-numerique-cest-quoi/

Ainsi, pourquoi laisser de côté le rôle de l'état, en plus du fait que cela correspond moins aux études en gestion ? Il y a plusieurs raisons qui motivent ce choix :

- Une surveillance est très difficile à estimer. En raison de son principe, celle-ci reste au plus possible confidentielle afin de maximiser son efficacité. Un système entièrement transparent est en effet très simple à contourner et perd donc toute son utilité. Il existe par ailleurs une réelle corrélation entre surveillance par un état et récolte de données par une organisation. Beaucoup de pays demandent une certaine coopération de la part de grandes entreprises et certains se permettent même de censurer différents services, en partie ou totalement. Cela varie bien entendu fortement en fonction des pays. A titre d'exemple, nous pouvons citer la Chine qui a réussi à mettre en place un réseau de surveillance extrêmement complexe, notamment à base de caméras et de reconnaissance faciale[5].

- Prendre en compte un tel rôle et s'en cacher devient fortement restrictif. Il est en effet presque impossible de contourner les obligations administratives au niveau de l'état, qui sont maintenant fortement numériques. Cela aborderait également des sujets beaucoup plus larges et encore moins dépendants de notre personne. Par exemple, et comme cité précédemment, il faudrait prendre en compte les caméras.

De plus, nous pourrions poser la question des raisons qui poussent un individu à se cacher du monde numérique. Comme dit précédemment, il s'agit du second sujet qui ne sera pas abordé dans ce travail, également pour plusieurs raisons :

- Tout d'abord, cette question est subjective. Il existe bien des éléments plus objectifs mais leur importance dépend des opinions de chacun. Dans un travail scientifique d'un ingénieur de gestion, la place d'un tel élément doit être minimisée.

- Au-delà même de la subjectivité, un tel sujet peut même être considéré comme philosophique. Comme déjà évoqué précédemment, il est important de cadrer un sujet en fonction de ses études. De cette manière, il n'est absolument pas pertinent et même légitime d'aborder ce sujet dans ce travail.

[5] DTP-AG (auteur anonyme), 2017, « La vidéosurveillance en Chine à des airs Big Brother », consulté le 23/08/2018, En ligne https://www.dtp-ag.com/surveillance-chine-big-brother/

- Finalement, une question si complexe pourrait également faire l'objet d'un mémoire à part entière. De nombreux auteurs s'y sont intéressés et en parlent largement. A titre d'exemple, nous pouvons citer Caroline Lancelot Miltgen (2011) qui aborde notamment quatre critères qui sont pris en compte par l'individu pour décider de s'il s'agit d'une intrusion dans sa vie privée ou non. Elle définit également le concept de protection de la vie privée, explique le lien entre individus et données, etc. Cela permet d'illustrer la richesse de cette question qui est finalement extrêmement vaste. Ainsi, en plus de l'absence de légitimité, le sujet deviendrait bien trop large.

Afin de répondre à notre question de recherche, nous passerons tout d'abord par une revue de la littérature. Nous y verrons les différents types de données ainsi qu'une évolution contextuelle de leur collecte. Pour les données divulguées de façon « involontaire » par leur propriétaire, nous aborderons les technologies de *tracking* et, pour celles de façon « volontaire », les éléments qui poussent un individu à les divulguer. Nous passerons par après aux conséquences de cette exploitation et nous nous pencherons sur les éléments de réponses que certains auteurs ont déjà apportés à notre question de recherche.

Nous dresserons ensuite un inventaire de nos traces numériques, divisé dans des catégories spécifiques aux smartphones, ordinateurs, navigation et services web, réseaux sociaux et éléments sortant du cadre du web. Le chapitre suivant décrira, pour ces mêmes catégories, les techniques, outils et habitudes à mettre en place pour limiter nos traces.

Afin de mettre cette section en application et d'apporter une réponse à notre question de recherche, nous nous pencherons sur une expérience ayant comme objectif de tester avec un individu s'il est possible de ne laisser aucune trace numérique durant une certaine période. Cette expérience a deux buts : vérifier de façon scientifique si des traces sont laissées pendant la durée du test, via l'utilisation du droit d'accès à nos données personnelles, et se rendre compte des difficultés d'un tel comportement. Les résultats de cette expérience seront ensuite confrontés à l'avis d'un expert du milieu et certaines décisions prises seront objectivées par une enquête. Enfin, nous aborderons les limites, difficultés et risques du comportement adopté durant l'expérience, tant pour la personne souhaitant minimiser ses traces que pour le monde des entreprises.

Revue de la littérature

Dans cette partie, nous allons résumer ce qui a déjà été apporté à notre sujet. Nous pourrons ensuite nous baser sur cela pour les chapitres suivants. Il est important de noter que la question des moyens pour se cacher du monde numérique a été très peu traitée par la communauté scientifique. Ceux-ci se penchent en effet plus sur différents aspects des données personnelles ou du *Big Data,* mais pas sur les techniques à mettre en place pour limiter son empreinte.

Dans ce chapitre, nous verrons tout d'abord le changement du contexte dans lequel se trouve le *Big Data*. Nous verrons notamment l'évolution des méthodes d'analyse et manières de représentations du monde, ou encore l'inversion du rapport à la mémorisation des données, celles-ci étant désormais enregistrées par défaut.

Nous nous concentrerons ensuite sur les données en elles-mêmes avec Bruce Schneier (2015) qui différencie 6 types de données en fonction de la façon dont elles sont générées. Nous nous pencherons également sur le travail de Louise Merzeau (2009) qui différencie « données » et « métadonnées » ou encore « identification » et « authentification ».

Pour les données fournies de façon involontaire, nous verrons une analyse complète sur les méthodes de *tracking* utilisées par les sites web, avec l'article « Les modèles d'affaires numériques sont-ils trop indiscrets ? Une analyse empirique » (Grazia Cecere et al., 2015).

Quant aux informations fournies de façon volontaire, nous verrons les raisons qui poussent un individu à dévoiler des données personnelles. Nous développerons l'important apport de Caroline Lancelot Miltgen (2011) et nous finirons sur le *privacy paradox*.

Ensuite, nous nous pencherons sur les conséquences, positives et négatives, de cette exploitation. Nous verrons notamment l'économie de la connaissance, le peu de protection juridique qui couvre la population ou encore l'asymétrie informationnelle.

Finalement, et notamment grâce à ce qui a déjà été développé, nous nous pencherons sur des auteurs ayant apportés des éléments de réponse à notre question de recherche, ce qui nous permettra de formuler une hypothèse.

Evolution de la collecte des données

Dans son article « La révolution Big data », Viktor Mayer-Schönberger (2014) aborde l'aspect historique du *Big Data*, à savoir les méthodes utilisées en ayant peu de données à disposition ainsi que la révolution des espaces de stockage, méthodes d'analyse et façons de représentations du monde. Louise Merzeau (2009) aborde elle aussi, dans « De la surveillance à la veille », un aspect historique au travers de l'inversion du rapport à la mémorisation des données.

Selon Viktor Mayer-Schönberger (2014), la collecte des données était, encore récemment, fastidieuse et loin d'être bon marché. L'objectif était donc d'obtenir un certain nombre d'informations avec un minimum de données. Pour cela, on appliquait la méthode de l'échantillonnage aléatoire, permettant d'utiliser une petite partie pour avoir une idée du tout. Cette solution montrait cependant ses limites et il était important de trouver d'autres solutions.

Aujourd'hui, nous avons à notre disposition d'énormes quantités de données, volume doublant tous les ans ou tous les deux ans. Celles numériques sont bien plus pratiques que les analogiques, que ce soit au niveau du stockage ou de l'analyse. Grâce à cela, et grâce aux changements technologiques engendrés, il fut par exemple possible pour Google de prédire la propagation de la grippe et ce en utilisant les requêtes envoyées à son moteur de recherche.

Cette évolution a permis au *Big Data* de mettre fin aux catégories *a priori* et laisse maintenant les ordinateurs déterminer les groupes. En effet, Amazon est rapidement passé de ceux fournis par des experts en marketing à des familles créées grâce à ses données, par exemple en utilisant les similarités entre produits vendus. Au moment où cet article a été écrit, en 2014, « le système de recommandation d'Amazon représenterait […] environ un tiers des revenus de l'entreprise ».

Louise Merzeau (2009), quant à elle, se penche sur l'inversion du rapport à la mémorisation des données. En effet, avec les technologies de stockage, l'oubli était le concept maitre jusque maintenant mais cela tend à changer. Elle cite Antoinette Rouvroy en expliquant que c'est désormais « par défaut [que] toute information (sonore, visuelle, textuelle) est enregistrée et conservée sous une forme digitale – l'oubli, nécessitant une action positive d'effacement des données, devenant de ce fait l'exception plutôt que la règle ». De plus, il est désormais plus couteux d'effacer l'information que de la garder par défaut.

En conclusion, la démocratisation des espaces de stockage a permis de passer de méthodes telles que l'échantillonnage aléatoire à de réelles analyses complexes rendant possible diverses prédictions. Ce changement fut également accompagné d'une révolution des méthodes d'analyse et de représentation du monde, comme par exemple avec la fin des catégories *a priori*. De plus, la norme est maintenant de conserver les données par défaut.

Les différents types de données

Dans son livre « Data and Goliath », Bruce Schneier (2015), un cryptologue et écrivain américain, différencie 6 types de données en fonction de la façon dont elles sont créées. Nous verrons l'explication de ces types et nous décrirons ensuite l'apport de Louise Merzeau (2009) qui distingue données et métadonnées ou encore identification et authentification.

Les 6 types de données de Bruce Schneier (2015) sont les suivants : celles de service, révélées, confiées, annexes, de comportement et dérivées.

- Les données de service : il s'agit des renseignements que nous fournissons par nous-même afin de pouvoir profiter d'un service. Par exemple, l'achat d'une carte SIM en Belgique implique ce type de des données étant donné qu'il est nécessaire de fournir sa carte d'identité. Actuellement, la majorité des sites proposant un service récolte ce genre de renseignements, demandes pouvant même parfois aller au-delà de ce qui est nécessaire pour profiter de ce service.

- Les données révélées : il s'agit d'informations, comme par exemple des messages ou des photos, divulguées volontairement sur des sites, blogs, etc. que nous possédons. L'aspect important ici est que nous détenons bien la plateforme, nous rendant ainsi maitre de ces données et du contrôle de celles-ci.

- Les données confiées : il s'agit des mêmes données que le point précédent mais, cette fois-ci, nous ne possédons pas la plateforme. De cette manière, nous ne contrôlons pas ce qui leur arrive. Ce type est beaucoup plus fréquent que le précédent car il inclut notamment tous les grands réseaux sociaux actuels.

- Les données annexes : dans ce groupe, nos données ne sont plus fournies directement par nous mais bel et bien par d'autres personnes. Comme exemple, nous pouvons citer le fait d'être identifié sur une photo Facebook, mentionné dans un tweet ou, pour sortir du cadre des réseaux sociaux, cité dans un article. Nous ne sommes cette fois-ci aucunement maitre de nos données et nous ne pouvons même pas choisir de les divulguer ou non. Cette catégorie comprend des données même plus larges que celles résultant d'une action volontaire d'une tierce personne car elle inclut aussi des actions involontaires, comme par exemple autoriser une application à avoir accès au répertoire de son téléphone.

- Les données de comportement : ce type est moins évident à estimer et surtout moins aisé à créer. Il a tendance à utiliser nos métadonnées afin d'essayer de cerner notre comportement, nos habitudes, ou notre façon d'agir. Cette catégorie comprend notre position géographique, combien de temps nous y sommes restés, avec qui nous étions, ou nous sommes allés ensuite, à quelle heure sommes-nous allés nous coucher, etc.

- Les données dérivées : ce dernier type est le seul ne comprenant pas de données qui ont été fournies par nous-même ou par des tiers. En effet, cette catégorie comporte tout ce qui a été inféré des 5 précédentes. Par exemple, une personne présentant certaines particularités est classée dans un certain cluster plutôt qu'un autre en fonction de ses caractéristiques. Dans cette catégorie, nous n'avons aucun contrôle direct sur nos données et nous ne pouvons agir que via les classes précédentes.

Louise Merzeau (2009) définit elle aussi différents types de données en se penchant sur la communication numérique. Cette communication est ainsi composée de deux niveaux, celui d'émission et de réception, à savoir des messages-cadres et des messages de contenu ou d'information. Dans le cadre d'une communication numérique, ces deux éléments peuvent désormais être séparés et analysés. De plus, l'information principale n'est désormais plus dans le message en tant que tel mais au niveau de l'information sur l'information, soit les métadonnées. L'individu a ainsi de plus en plus de difficultés à contrôler ses traces car « l'information utile s'est déplacée de la surface lisible des messages aux couches internes, moins déchiffrables ».

Un autre apport de l'auteure sur ce sujet est sa distinction entre l'identification, à savoir le fait de mettre un nom sur une trace, et l'authentification, identifier une trace avec l'état civil de son propriétaire. Il n'est souvent pas nécessaire d'authentifier une trace, l'identification via de fausses informations faisant l'affaire. Nous pouvons ainsi très bien laisser des données identifiées mais non-authentifiées.

En résumé, selon Bruce Schneier (2015), il existe 6 grands types de données, principalement différenciés par la façon dont celles-ci sont créées. Les données de services sont celles fournie par un individu afin de pouvoir profiter d'un service, celles révélées sont divulguées sur un site contrôlé par l'utilisateur, a contrario des données confiées où la personne n'est pas la propriétaire de la plateforme. Les données annexes ne sont cette fois plus divulguées par l'utilisateur lui-même mais par des tiers et les données de comportement résident

principalement en des métadonnées telles que la position géographique, combien de temps nous sommes restés à un endroit, etc. Enfin, les données dérivées sont des éléments déduits de celles présentes dans les cinq catégories précédentes. Louise Merzeau (2009) distingue les données des métadonnées et déclare que ces dernières contiennent désormais l'information principale, rendant compliqué un contrôle de ses traces. Elle différencie également l'identification de l'authentification.

Les technologies de *tracking*

Dans leur article « Les modèles d'affaires numériques sont-ils trop indiscrets ? Une analyse empirique », Grazia Cecere, Fabrice Le Guel et Fabrice Rochelandet (2015) analysent principalement l'utilisation des traceurs par les sites web. Nous définirons tout d'abord le concept de « traceur » avant de faire un rappel historique sur l'évolution de cette technologie. Nous verrons ensuite les principales conclusions de leur expérience, à savoir dans quelle mesures les plus grands sites utilisent des technologies de *tracking*, et nous nous pencherons par après sur l'utilisation qui est réalisée avec ces données. Celles-ci peuvent en effet profiter aux sites ou même aux utilisateurs, toujours avec des limites. Finalement, nous verrons qu'il est important de ne pas réaliser une collecte et une utilisation trop intensive des données personnelles.

Tout d'abord, il faut définir le concept de « traceur » : il s'agit de petits fichiers informatiques générés automatiquement, souvent à l'insu de l'individu, qui enregistrent ses traces numériques. Ces éléments créent ensuite des données individuelles qui peuvent elles-mêmes devenir des données personnelles, lorsque l'on arrive à les relier à une identité. Nous pouvons d'ailleurs faire un lien avec Louise Merzeau (2009) et ses concepts d'identification et authentification.

Historiquement, les technologies de *tracking* évoluent fortement. Aux débuts d'internet, l'adresse IP de l'ordinateur est utilisée pour identifier un individu. Cependant, il existe rapidement des moyens de cacher cette adresse, rendant cette technologie dépassée. D'autres solutions sont ainsi développées, comme l'ajout d'un identifiant unique dans les pages HTML ou dans l'URL. Enfin, la technologie des cookies voit le jour. Ceux-ci peuvent provenir des sites visités ou de parties tierces, exister le temps d'une session ou être effacés au bout d'une durée déterminée.

De nouvelles méthodes voient encore le jour, notamment depuis l'apparition du HTML5 dans les navigateurs internet. En effet, les sites web visités peuvent désormais utiliser l'ordinateur de leurs visiteurs pour stocker d'importantes quantités d'informations. Une autre technologie également fortement utilisée est le « *Device Finger Prints* », concept permettant de reconnaitre un internaute grâce aux particularités de son ordinateur, à savoir au travers des caractéristiques de son système d'exploitation, du navigateur web, de la taille de l'écran, etc.

Afin d'aider les sites web dans cette tâche, beaucoup d'acteurs apportent leur expertise afin de profiler les visiteurs, permettant la diffusion de publicités ciblées ou même la vente de ces données dans des enchères en temps réel. Il existe en effet un « second marché » des données personnelles exploité par de nombreux acteurs tiers inconnus des internautes.

Les auteurs consultent ensuite les 200 sites les plus visités en France afin d'observer dans quelle mesure ceux-ci utilisent des outils de *tracking*. Il est important de noter que cette étude est réalisée en 2011, la situation ayant fortement évolué depuis. Les principales conclusions sont que « les firmes qui utilisent le plus les traceurs (cookies, images invisibles, etc.) sont les sites internet d'information et des journaux ainsi que les sites de communication institutionnelle des entreprises. En revanche, les sites de réseautage social et de rencontre en ligne sont nettement moins dépendants de ces techniques et collectent directement les données personnelles par le biais des déclarations des individus au moment de leur inscription et surtout lors de leur utilisation de ces services ».

Une fois les données sur leurs clients récoltées, celles-ci peuvent être utilisées de façons très différentes : « identifier ces derniers, communiquer ou interagir avec eux de façon plus ou moins ciblée, différencier leurs prix ou la qualité de leurs offres en fonction des profils des consommateurs identifiés (marché du crédit bancaire par exemple), animer une communauté en ligne, faciliter l'appariement entre les utilisateurs de leurs services (services de réseautage social ou de rencontre en ligne par exemple), revendre ces données à d'autres firmes ».

Cette collecte peut bénéficier principalement à ses exploitants, comme c'est par exemple le cas pour la vente des données, ou profiter mutuellement à ceux-ci et aux utilisateurs. En effet, les entreprises peuvent ainsi offrir des services spécifiques et personnalisés ou encore profiter de revenus tirés de la publicité ciblée. Les consommateurs, quant à eux, bénéficient d'un service gratuit et « retirent une valeur d'usage de contenus ». Cependant, il faut garder à l'esprit que ces utilisateurs peuvent souffrir d'externalités négatives, comme cela serait le cas pour une discrimination des prix.

Il faut également faire attention à ne pas réaliser une collecte et une utilisation trop intenses des données personnelles, ce qui pourrait mener à une perte de confiance des utilisateurs et à des réactions négatives. Ceux-ci pourraient en effet bloquer les publicités ciblées via des applications ou extensions. Afin du nuancer cela, ces chercheurs ont prouvé que « les sites les plus intrusifs étaient dans les faits les plus performants ».

Il faut finalement être conscient que les internautes sont désormais dans une « ère post-cookies » grâce aux nouveaux types de traceurs mais aussi grâce aux réseaux sociaux. Les individus fournissent de manière volontaire leurs données en échange de services et contenus. De simples outils antitraceurs classiques ne suffisent donc plus.

En conclusion, les technologies de *tracking* ont fortement évolué, allant de l'adresse IP au « *Device Finger Prints* ». Cette évolution a même permis l'apparition de nouveaux acteurs spécialisés dans le profilage des internautes et dans la vente de leurs données. Nous savons également que les sites utilisant le plus les technologies de *tracking* sont « les sites internet d'information et des journaux ainsi que les sites de communication institutionnelle des entreprises ». Cependant, les réseaux sociaux n'en réalisent pas une grande utilisation, leurs adhérents fournissant par eux-mêmes de nombreuses informations personnelles. De telles données servent ensuite à diverses tâches et peuvent bénéficier à toutes les parties, même si c'est plus souvent le cas pour les exploitants. Enfin, il faut faire attention à ne pas réaliser une collecte et une utilisation trop intensive et engendrer ainsi des réactions négatives.

Eléments incitant un individu à divulguer ses données

Dans son article « Vie privée et marketing. Étude de la décision de fournir des données personnelles dans un cadre commercial », Caroline Lancelot Miltgen (2011) se penche sur le concept de protection de la vie privée et, plus particulièrement, sur ce qui incite un individu à dévoiler des données personnelles ou non. Nous verrons tout d'abord la définition de la protection de la vie privée, nous passerons ensuite aux éléments qui poussent une personne à divulguer ses données, dont notamment les quatre critères pris en compte, et nous finirons sur le *privacy paradox*.

De plus en plus d'individus ressentent une réelle intrusion dans leur vie privée. En effet, plusieurs travaux et sondages montrent qu'il y a aujourd'hui beaucoup de personnes qui perçoivent l'utilisation de leurs données comme telle.

L'auteure définit ensuite le concept de « protection de la vie privée » au travers de plusieurs dimensions principales : le droit à l'information, le droit au consentement, le contrôle sur l'utilisation ultérieure des données et sur les intrusions non désirées. Cette définition permet de se rendre compte de l'étendue du concept et donc du travail à réaliser.

Elle se penche ensuite sur ce qui pousse une personne à dévoiler ses données et remarque qu'il s'agit d'un réel calcul couts/bénéfices de sa part. Afin d'approfondir le sujet, elle réalise une série d'interviews. Elle observe tout d'abord que tous les individus n'ont pas la même perception quant à leurs données personnelles et à leur vie privée mais aussi que ceux-ci présentent des attitudes et connaissances fortement différentes face aux pratiques de collecte des données.

Ainsi, grâce à ses entretiens, elle arrive à repérer quatre critères qui sont pris en compte par une personne pour savoir si elle va dévoiler ses données ou non :

- La sensibilité perçue des informations demandées : même si la sensibilité est subjective et dépend donc de chaque individu, la nature des données demandées joue un rôle important.
- La confidentialité perçue des données : il s'agit de la confiance qu'a l'individu envers l'entreprise pour garder ses données confidentielles.
- La pertinence perçue de la sollicitation : les consommateurs s'interrogent en effet souvent sur ce qui pousse les entreprises à demander telle ou telle information.

- La valeur perçue de l'échange : il s'agit du calcul couts/bénéfices réalisé par l'utilisateur. Il faut cependant garder à l'esprit qu'il s'agit d'un sentiment et non d'un calcul objectif et absolu.

L'auteure s'attarde finalement sur le *privacy paradox*, la double volonté de vouloir être connu et reconnu mais, en même temps, de chercher à limiter les intrusions dans sa vie privée. Ses entretiens lui ont notamment permis de voir que beaucoup d'individus oublient purement et simplement leurs préoccupations concernant leur vie privée quand ils sont en ligne.

Louise Merzeau (2009) définit elle aussi le *privacy paradox,* comme le fait de livrer de plus en plus d'informations personnelles tout en redoutant d'être surveillé et suivi. Ce concept s'applique en effet à beaucoup de personnes, surtout depuis plusieurs années, étant donné que les utilisateurs exigent une certaine fluidité, notamment dans les interfaces web. Il faut que celles-ci se souviennent de tous les états précédents afin de garantir un service de qualité, les forçant ainsi à garder cette mémoire.

Selon Fabrice Rochelandet (2010), ce paradoxe et cette « surexploitation de soi, par exemple sur les réseaux numériques » rendent compliqué un réel contrôle de nos traces numériques.

En conclusion, de plus en plus d'individus ressentent une réelle intrusion dans leur vie privée lorsque des sites essayent de récolter des données les concernant. Ceux-ci n'ont également pas le même comportement et les mêmes connaissances face à la récolte de leurs données. Ils jugent cependant acceptable de les fournir si « 1) les renseignements demandés sont perçus comme peu sensibles ; 2) ils peuvent contrôler l'utilisation future des données (respect de leur confidentialité) ; 3) les informations demandées sont jugées pertinentes pour la transaction en cours ; 4) ils ont le sentiment qu'ils ont plus à gagner qu'à perdre en fournissant ces données (valeur perçue positive) ». Enfin, on remarque que beaucoup de personnes oublient leurs préoccupations concernant leur vie privée lorsque qu'elles sont en ligne et qu'elles exigent une fluidité toujours croissante, observation expliquant le *privacy paradox.* Un tel phénomène complique d'ailleurs un contrôle de ses traces numériques.

Conséquences de l'exploitation des données

Dans cette partie, nous allons décrire les conséquences, positives et négatives, de cette exploitation. Louise Merzeau (2009) et Viktor Mayer-Schönberger (2014) développent un aspect plus économique en parlant d'une « économie de la connaissance », d'e-réputation et de nos données comme une nouvelle ressource économique. Nous verrons ensuite des aspects négatifs tels que le peu de protection juridique qui couvre la population, la capacité à réaliser de plus en plus de prévisions ou encore « l'asymétrie informationnelle ».

Selon Louise Merzeau (2009), toutes ces données de moins en moins contrôlables et toujours plus nombreuses permettent la création d'une « économie de la connaissance » dans laquelle s'engouffrent les états et les entreprises. Celle-ci sert notamment à réaliser des prévisions permettant de réduire les incertitudes en créant des « bases de données de nos intentions ». L'objectif est donc d'anticiper les agissements plutôt que de les contraindre.

De plus, étant donné l'importance grandissante de ces traces, un réel marché de l'e-réputation voit le jour et compte divers spécialistes, entreprises et services. Si nous souhaitons profiter de ces services, il existe cependant beaucoup d'outils à prendre soi-même en main sans passer par un acteur professionnel.

Selon Viktor Mayer-Schönberger (2014), l'enjeu pour les sociétés occidentales est de prévoir l'avenir afin de diminuer le risque. Grâce aux importants volumes de données, il est maintenant possible d'avoir une vision exhaustive du monde.

Il se penche également sur nos informations qui deviennent selon lui une nouvelle ressource économique. Contrairement à beaucoup de ressources matérielles, un usage n'est pas forcément unique et la valeur n'est pas détruite ou dégradée lors de celui-ci. La chute des couts de stockage et d'analyse a ainsi permis la création de nouveaux *business models*, profitant aux grandes entreprises mais également aux plus petites. Ainsi, une start-up américaine, Inrix, utilise les données de position et de vitesse de ses utilisateurs afin de les aider à éviter les embouteillages. De telles analyses lui ont même permis de découvrir des corrélations entre ses données de circulation aux alentours des centres commerciaux et les revenus des commerces qui y sont implantés. L'entreprise a ainsi créé un modèle permettant de prédire les revenus de ces magasins, ressources ayant permis la mise en place de partenariats avec des fonds d'investissement. Le *Big Data* et les données de façon générale sont, de cette manière, un enjeu majeur pour tous les types d'entreprises ou même pour les gouvernements.

L'auteur évoque cependant plusieurs côtés négatifs comme, par exemple, le peu de protection juridique qui couvre la population. Il critique également cette capacité à réaliser de plus en plus de prévisions, éléments pouvant avoir des conséquences plus importantes que l'acte prédit en lui-même. Pour illustrer cela, il demande ce qu'il se passerait si nous étions capables de prévoir si une personne sera une bonne conductrice ou non. Est-ce que des assureurs l'accepteraient malgré une prévision négative ?

Finalement, il aborde le sujet de l'avenir et notamment le débat de la gouvernance de l'internet. L'enjeu ne sera en effet plus internet au sens strict mais bien le contrôle et l'exploitation des données. De plus, le *Big Data* est un domaine toujours en évolution, il « va faire évoluer la façon dont nous donnons du sens au monde, et transformer ce que nous considérons comme étant source de valeur économique ».

Une autre critique est adressée par Fabrice Rochelandet (2010), selon qui nous sommes passé d'une simple collecte, comme c'était par exemple le cas pour les états, à une véritable exploitation économique. L'auteur insiste beaucoup sur les conséquences négatives et notamment sur « l'asymétrie informationnelle », soit le fait que les industriels possèdent plus d'informations que les individus sur leurs données. Une autre externalité négative est également la discrimination des prix, pratique existante étant donné qu'Amazon fut condamné pour cela en 2000.

Il déclare ensuite que la barrière principale empêchant une entreprise d'utiliser les données dont elle dispose à des fins outrepassant ses engagements envers ses clients est « le risque de sanction légale » et « l'effet de réputation ». L'entreprise agit ainsi uniquement dans son propre intérêt et non celui de ses clients.

Une critique similaire est adressée plus particulièrement au géant américain Google par Barbara Cassin (2009). Elle dénonce son côté *Big Brother et* critique l'algorithme *PageRank* du moteur de recherche. Elle aborde la place de la publicité dans les revenus du géant, son rapport à l'opinion et sa nature, à savoir renseigner et non éduquer. Comme il s'agit d'une critique envers une entreprise précise et non d'une généralité, seuls les points principaux de l'article sont cités. Pour une explication détaillée, les notes de lecture complètes sont disponibles en annexe 1.

Ainsi, toutes ces données ont permis la création d'une « économie de la connaissance » et d'un commerce de l'e-réputation dont dépendent désormais beaucoup d'entreprises. Les informations sont une nouvelle ressource économique dont l'usage n'est pas forcément unique et la valeur n'est pas détruite ou dégradée lors de celui-ci. Grâce à celles-ci, de nouvelles sources de revenus et même de nouveaux *business models* ont vu le jour, profitant aux grandes entreprises mais également aux plus petites. Cependant, il existe souvent un réel manque de protection juridique de la population. De plus, les nouvelles méthodes d'exploitation de ces données ou encore « l'asymétrie informationnelle » peuvent engendrer des dérives.

Est-il possible de se cacher du monde numérique ?

Certains auteurs ont réalisé des apports très intéressants à notre question de recherche. Nous verrons une auteure qui affirme que « on ne peut pas ne pas laisser de traces » ainsi que l'importance de la visibilité en ligne pour le développement d'un individu.

Louise Merzeau (2009) affirme qu'on ne peut pas ne pas communiquer et, surtout, que « on ne peut pas ne pas laisser de traces ». Il n'y a en effet plus de « comportement zéro » depuis que l'information est désormais la norme. L'utilisateur laisse ainsi d'innombrables empruntes derrière lui mais est également victime de toutes celles laissées par des tierces personnes. Ces traces peuvent être très larges et incluent notamment le fait de parler de quelqu'un, l'identifier quelque part, le citer dans un article ou encore se lier d'une quelconque façon à un de ses contenus. Ces informations qui autrefois étaient inutilisables sont désormais exploitables grâces aux nouvelles technologies, comme nous l'avons également vu dans l'évolution des méthodes de *tracking* (Grazia Cecere et al., 2015).

Fabrice Rochelandet (2010) se penche également sur les limites d'un contrôle de ses traces. Il emprunte le modèle de George A. Akerlof et Rachel E. Kranton, à savoir la théorie de l'utilité de l'identité sociale : « tout individu choisit d'appartenir à une catégorie sociale donnée, afin de se situer et d'être situé ». Ainsi, il a été prouvé que la visibilité en ligne, principalement sur les réseaux sociaux, est importante pour le développement d'une personne. Ces sites sont donc presque indispensables même s'ils sont également les responsables d'une grande part de nos traces, élément confirmé par Grazia Cecere et al. (2015). Il ne faut également pas oublier le rôle du *privacy paradox* qui complique encore plus le contrôle de ses traces (Fabrice Rochelandet, 2010).

En résumé, un individu laisse d'innombrables traces derrière lui, notamment à cause des conséquences du *privacy paradox*, et est également victime de toutes celles laissées par des tierces personnes. De plus, les réseaux sociaux jouent désormais un rôle important dans le développement personnel et sont presque indispensables alors qu'ils génèrent d'innombrables traces numériques. Ces éléments nous incitent à poser l'hypothèse qu'il n'est, dans notre société actuelle, pas possible de se cacher du monde numérique.

Comme nous avons pu le voir dans cette revue de la littérature, la démocratisation des espaces de stockage a permis de passer de méthodes telles que l'échantillonnage aléatoire à de réelles analyses complexes permettant de réaliser des prédictions. Ce changement fut également accompagné d'une révolution des méthodes d'analyse et de représentation du monde, comme par exemple avec la fin des catégories *a priori*. De plus, la norme est maintenant de conserver les données par défaut.

Bruce Schneier (2015) distingue, quant à lui, 6 grands types de données. Celles-ci se placent dans une catégorie ou une autre en fonction du contexte de leur création. Par exemple, certaines dépendent de l'utilisateur lui-même alors que d'autres non, certaines sont également différentes en fonction de qui possède la plateforme. Louise Merzeau (2009) parle de métadonnées qui contiennent désormais l'information principale et différencie l'identification de l'authentification.

Grazia Cecere, Fabrice Le Guel et Fabrice Rochelandet (2015) présentent l'évolution des technologies de *tracking*. Ils réalisent une étude sur les sites les plus visités et concluent que les sites utilisant le plus ces traceurs sont « les sites internet d'information et des journaux ainsi que les sites de communication institutionnelle des entreprises ». Ils expliquent cependant que les réseaux sociaux n'ont pas autant besoin de ces technologies car que leurs utilisateurs fournissent eux-mêmes de nombreuses informations personnelles.

Caroline Lancelot Miltgen (2011) établit que les individus jugent acceptable de fournir leurs données si « 1) les renseignements demandés sont perçus comme peu sensibles ; 2) ils peuvent contrôler l'utilisation future des données (respect de leur confidentialité) ; 3) les informations demandées sont jugées pertinentes pour la transaction en cours ; 4) ils ont le sentiment qu'ils ont plus à gagner qu'à perdre en fournissant ces données (valeur perçue positive) ».

En conséquence, toutes ces données ont permis la création d'une « économie de la connaissance », l'apparition de nouvelles sources de revenus et même de nouveaux *business models.* Il existe cependant de nombreux aspects négatifs comme le manque de protection juridique couvrant la population, cette nouvelle capacité de prédiction et « l'asymétrie informationnelle ».

Finalement, Louise Merzeau (2009) établit que « on ne peut pas ne pas laisser de traces » car l'utilisateur est victime des traces le concernant laissées par des tierces personnes. De plus, les réseaux sociaux jouent désormais un rôle important dans le développement d'un individu et sont presque indispensables, même s'ils sont responsables d'une grande partie de nos traces. Nous pouvons ainsi poser l'hypothèse qu'il n'est pas possible de se cacher du monde numérique.

Méthodologie

Nous souhaitons vérifier l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique, et donc qu'il n'est pas possible de ne pas laisser de traces. Pour cela, nous listons l'ensemble des traces numériques que nous laissons et nous utilisons cet inventaire pour trouver des solutions permettant de réduire notre emprunte numérique. Ces deux premiers éléments plus théoriques nous conduisent à la pratique via la réalisation d'une expérience dont les résultats obtenus sont ensuite confrontés à l'avis d'un expert. Cette expérimentation nous permet par ailleurs de réaliser une dernière partie de nouveau plus théorique comprenant les limites, difficultés et risques du comportement adopté par un individu souhaitant minimiser ses traces.

Ainsi, nous dressons dans un premier temps l'inventaire des traces numériques que nous pouvons laisser. Comme un tel inventaire n'existe pas, il est nécessaire de l'élaborer. Pour cela, nous nous aidons d'une importante lecture d'articles scientifiques, d'articles de presse, de sites spécialisés et d'expériences personnelles. En construisant ce chapitre, nous pouvons observer que certaines traces sont communes sur certains points et des catégories sont donc réalisées en fonction de leur nature, afin d'ordonner cet inventaire et de faciliter la recherche de solutions.

Nous passons ensuite par le développement de solutions pour limiter les traces que nous laissons. Pour ce chapitre, l'inventaire précédemment écrit est utilisé et, pour chaque trace, nous cherchons une réponse dans la littérature. Après avoir parcouru l'inventaire, en passant en revue les différents types de données de Bruce Schneier (2015), nous nous rendons compte que nous ne couvrons pas les données dérivées. La seule solution pour limiter ce type de données étant de supprimer ses traces actuelles, nous réalisons une partie dédiée à cela.

Afin de pouvoir tester concrètement les solutions développées précédemment, nous réalisons une expérience dont l'objectif est d'essayer de se cacher du monde numérique pendant au minimum une semaine. Celle-ci a plusieurs aspirations : obtenir une preuve concrète qu'il n'est pas possible de ne pas laisser de traces et mieux se rendre compte des limites, difficultés et risques d'un tel comportement. Afin de sélectionner les mesures à appliquer pour limiter ses traces durant cette expérimentation, nous utilisons le chapitre concernant les solutions à mettre en place. L'ensemble de ces mesures est par ailleurs détaillé en annexe 3 avec le carnet de bord. La durée de cette expérience est, quant-à-elle, variable. L'objectif est de la faire durer au minimum une semaine complète et de l'arrêter quand aucun élément nouveau n'est à signaler dans le carnet de bord.

Par après, nous faisons valoir notre droit d'accès à nos données personnelles auprès de nombreuses entreprises afin d'observer si certaines ont généré de nouvelles données durant notre expérimentation, ce qui validerait l'hypothèse. Pour savoir auprès de qui réaliser ces demandes, nous reprenons l'inventaire des traces numériques afin d'observer quelles catégories s'appliquent à notre situation. Il est également important de noter qu'il est parfois nécessaire de réaliser de multiples demandes au sein d'une même catégorie, par exemple en cas de possession de comptes sur plusieurs réseaux sociaux.

La méthode appliquée pour réaliser cette expérience présente cependant des limites. En effet, l'aléatoire peut jouer un rôle important, d'autant plus que cet exercice n'est exécuté qu'avec un seul individu. En effet, nous sommes dépendants des traces laissées spécifiquement par cette personne mais également des évènements aléatoires, comme par exemple la réception d'un email. Pour pallier cela, nous confrontons l'analyse des résultats de cette expérience à l'avis d'un expert du milieu. Nous réalisons ainsi une interview de Bob de Schutter, docteur à l'université de Gand travaillant actuellement en tant que chef du Data Engineering dans une entreprise de e-commerce. Il faut également garder à l'esprit qu'une telle approche présente également ses limites car qu'il s'agit d'un avis personnel, aussi neutre soit-il.

De plus, afin d'objectiver certains aspects de ce travail et de rationaliser les mesures prises dans l'expérience, une courte enquête est réalisée. Celle-ci est créée en ligne avec l'outil LimeSurvey et compte 241 répondants. L'avantage majeur de cette approche est de pouvoir faire intervenir l'avis de tiers ayant différents profils socioculturels. Gardons cependant à l'esprit qu'une telle enquête peut être biaisée malgré l'anonymat, et ce par exemple avec la désirabilité sociale. Un tel effet pourrait par exemple se caractériser au travers d'une personne déclarant pouvoir se passer sans difficulté d'un smartphone pendant 10 jours car il n'est, dans son environnement social, pas bien vu d'avouer être en partie dépendant de son téléphone bien que ce soit le cas. Il est donc pertinent d'utiliser une telle approche en complémentarité de notre expérience et de l'interview d'expert.

Finalement, nous arrivons à la dernière partie du corps de ce travail, soit les limites, difficultés et risques. La logique de cette partie est de suivre le processus complet de causalité des outils développés et de l'expérimentation. Nous nous penchons en effet déjà sur la cause, soit la volonté de ne pas laisser de traces listées dans notre inventaire, et il nous reste à développer les conséquences, soit les risques. Ce chapitre comporte le développement des risques pour la personne adoptant ce comportement et pour les entreprises. Ces deux éléments sont alimentés par la littérature et le premier est également en partie basé sur l'expérience réalisée.

Ainsi, la méthodologie appliquée débute par l'établissement d'un inventaire des traces numériques que nous laissons. Celui-ci est ensuite repris pour développer des solutions permettant de réduire nos traces, chapitre également mis en application au travers d'une expérience. Notre analyse des résultats de cette dernière est confrontée à l'avis d'un expert du milieu et une enquête est également réalisée. Enfin, nous développons les limites, difficultés et risques du comportement d'une personne souhaitant minimiser ses traces.

Inventaire des traces numériques que nous laissons

Dans ce chapitre, nous tacherons de lister toutes nos traces numériques, divisées dans les catégories suivantes : les traces spécifiques aux smartphones, les traces spécifiques aux ordinateurs, la navigation web et certains services populaires. Notons également que ces catégories incluent les objets connectés, comme par exemple les *smartwatch*. Ensuite, nous nous concentrerons sur les réseaux sociaux et finalement sur une catégorie sortant du cadre du web comprenant les informations bancaires, les cartes de fidélité et d'autres services nécessitant la divulgation de données. Il est également important de préciser que l'objectif est de dresser une liste complète pour la majorité des personnes mais que chaque individu laisse des traces différentes.

Smartphones

Presque indispensables de nos jours, nos smartphones sont une source très importante de nos traces numériques. En juillet 2018, ils représentent par ailleurs 53% des parts de marché lorsque nous les comparons aux ordinateurs, 43%, et aux tablettes, 4%[6]. Dans cette partie, nous allons nous attarder sur les traces spécifiques aux smartphones en laissant de côté la navigation et certains services web dont les réseaux sociaux. Nous verrons ainsi le smartphone en lui-même avec son système d'exploitation (OS), l'opérateur téléphonique, la localisation GPS et un point sur les applications. Il est à noter que cette partie comprend également les traces générées par les tablettes tant celles-ci ont un fonctionnement similaire.

- Smartphone en tant que tel : pour celui-ci, nous pouvons donner différentes informations au constructeur. Nous laissons des données de service, comme par exemple avec la création d'un compte Samsung ou Apple où nous fournissons beaucoup d'informations personnelles. Un tel compte donne la possibilité à l'entreprise de réaliser un lien entre notre profil et toutes nos autres données, ce qui permet d'en créer des identifiées voire même des authentifiées (Merzeau, 2009).

- Système d'exploitation du smartphone : nous laissons voir l'utilisation que nous réalisons du smartphone, soit la quantité de batterie que nous utilisons, quand nous réalisons nos recharges, combien de temps le téléphone a été en veille, combien de temps et à quel moment est-il en utilisation, quelles applications ont été téléchargées et utilisées, etc. Il serait très compliqué de lister l'ensemble des données que nous laissons étant donné que cela pourrait même concerner la totalité de nos options dans les réglages et préférences.

- Opérateur téléphonique : initialement, le rôle d'un téléphone était simplement de pouvoir passer des appels et, par la suite, d'envoyer des messages. Ainsi, chaque smartphone est équipé d'une carte SIM et laisse via celle-ci différentes traces. Nous pouvons tout d'abord penser à l'exemple utilisé précédemment pour les données de service. Lors de la souscription d'un abonnement, plusieurs informations doivent être laissées à l'opérateur. De plus, même sans abonnement, il est désormais impossible d'acheter une simple carte SIM sans présenter de pièce d'identité.

[6] StatCounter Global Stats, 2018, « Desktop vs Mobile vs Tablet Market Share Worldwide », consulté le 12/08/2018, En ligne http://gs.statcounter.com/platform-market-share/desktop-mobile-tablet

Ensuite, nous sommes dépendants des antennes relais. Celles-ci permettent ainsi de nous situer dans une zone géographique, en Belgique ou à l'étranger, et de savoir combien de temps nous sommes restés dans ces zones.

Nous laissons également diverses informations quant à notre utilisation de ces services. Nous pouvons par exemple citer la quantité de données mobiles utilisées, le nombre de SMS envoyés, à quel numéro et à quelle heure, et idem pour les appels.

Toutes ces données sont notamment utilisées pour réaliser du clustering et du profilage[7]. Il n'est en effet pas rare de recevoir des offres spéciales correspondant à notre utilisation.

- Localisation GPS : à chaque fois que le paramètre est activé, beaucoup de données sont générées. De base très simples, celles-ci peuvent cependant évoluer en données dérivées et devenir très complexes. Ainsi, l'entreprise Apple réussit à déduire qu'un endroit est notre logement car il s'agit du lieu dans lequel nous restons souvent le soir[8]. Avec la même logique, il est aussi facile de déduire le travail, le supermarché préféré, les endroits dans lesquels nous aimons aller, etc. Cette localisation est également utilisée dans les métadonnées d'autres systèmes comme, par exemple, les photos que nous prenons ou encore d'où nous postons quelque chose sur un réseau social. Ces informations permettent ainsi aux entreprises d'enrichir un profil afin de fournir un service plus personnalisé. Cela a donc des intérêts pour celles-ci, qui pourront proposer des produits complémentaires ou nouveaux, mais cela profite également à l'utilisateur qui peux ainsi bénéficier d'un service d'une qualité accrue.
- Applications : elles sont responsables d'une importante part de nos traces. Contrairement aux autres parties ayant pour objectif d'être plus générales, celle-ci est relativement plus personnelle et a pour but d'illustrer que nous laissons d'innombrables traces au travers de nos applications. La caractéristique majeure est que ces outils ont l'obligation d'être clairs sur les fonctionnalités utilisées et ils doivent même demander l'autorisation à l'utilisateur. Notons que nous n'abordons pas les applications basiques, comme par exemple « Messages » ou « Téléphone », étant donné qu'elles sont

[7] Proximus, 2019, « Mentions légales pour les clients privés et professionnels », consulté le 26/05/2019, En ligne https://www.proximus.be/fr/id_cr_warnland/particuliers/produits/r-orphans/informations-legales.html#/tab2

[8] Me and my Shadow (auteur anonyme), 2017, « Location tracking », consulté le 03/08/2018, En ligne https://myshadow.org/location-tracking

comprises dans le système d'exploitation. Ainsi, voici certaines applications pouvant récolter des données :

- Play Store : application indispensable pour un possesseur d'appareil Android, « Google Play Store » est tout simplement une application installée par défaut et qui permet d'installer toutes les autres. Cette plateforme connait ainsi toutes les applications qui ont été installées, l'éventuelle évaluation que nous leur avons accordée, le temps que nous les avons laissées sur l'appareil avant de les désinstaller, etc. Toutes ces informations permettent encore d'enrichir un profil qui, de cette manière, rend possible la création d'un système de recommandation afin de suggérer d'autres produits. Il est en effet important pour l'entreprise que l'utilisateur passe du temps sur des applications qu'il apprécie étant donné que la majorité de celles qui sont gratuites passent des publicités via Google qui récupère une commission[9]. Pour celles qui sont payantes ou qui proposent des microtransactions, le géant américain récupère également une commission lors de chaque achat[10]. Il va sans dire que cela fonctionne exactement de la même façon pour son concurrent Apple ou pour d'autres applications proposant des services similaires.
- Camera : même s'il s'agit d'une application plus basique dépendante du système d'exploitation, l'appareil photo du smartphone laisse un certain nombre de traces. En effet, en plus des informations quant à la fréquence ou à l'heure de prise de photos, chaque image contient des métadonnées. Par exemple, si le GPS est activé, la localisation est gardée en mémoire. De plus, l'application est même capable de reconnaitre ce que contient une image, comme par exemple de la nourriture ou un paysage.
- Agenda : il est difficile d'estimer dans quelle mesure nos données de calendrier sont utilisées. Il est cependant clair que l'application est capable de reconnaitre ce qui est écrit en intitulé d'un élément de l'agenda. Par exemple, elle adapte

[9] Google, 2018, « Google AdMob - Mobile App Monetization & In App Advertising », consulté le 13/08/2018, En ligne https://www.google.com/admob/

[10] Google Play Console Help, 2018, « Transaction fees », consulté le 22/08/2018, En ligne https://support.google.com/googleplay/android-developer/answer/112622?hl=en

automatiquement l'image illustrative d'un évènement en fonction de son nom, comme nous pouvons le voir sur la figure 1.

Figure 1 : Evènement d'un agenda Google

- Application « National » : cette application permet de planifier des trajets en transports en commun dans toute la Belgique. Des données comme l'historique des trajets, les lieux favoris, etc. pourraient donc être récupérées. Cependant, aucun compte n'est nécessaire, rendant compliqué un éventuel lien entre des traces et un profil. Les informations personnelles ne doivent être fournies que lors d'un achat direct de titres de transport.

- Shazam : cette application permet de reconnaitre une musique via le microphone du téléphone afin d'obtenir des informations la concernant. Il ne faut pas créer de compte mais des données quant aux musiques écoutées peuvent être récupérées. Il est également intéressant de noter que Shazam appartient à Apple et pourrait donc être combiné à d'autres services.

- MyProximus : cette dernière application est liée à l'opérateur téléphonique. Celle-ci incite même fortement l'utilisateur à créer un compte et peut de cette manière plus facilement lier les données générées au profil. Elle permet également de connaitre plus facilement les applications préférées, notamment via la consommation spécifique de données mobiles.

En résumé, nous pouvons laisser des informations au constructeur du smartphone lors d'une éventuelle inscription. Le système d'exploitation, quant à lui, récolte une très grande quantité de données dans des domaines très larges. Nous laissons ensuite plusieurs traces à notre opérateur téléphonique ou encore via notre localisation GPS. Enfin, les applications jouent un rôle capital dans nos traces numériques. Nous pouvons notamment citer le Play Store pour qui il est très important d'avoir un excellent système de recommandation.

Ordinateurs

Presque aussi utilisés que nos smartphones et tout aussi indispensables, les ordinateurs sont également responsables d'une grande part de nos traces numériques. Dans cette partie, nous allons voir les traces spécifiques aux ordinateurs en laissant de côté les réseaux sociaux et les services web. Nous nous pencherons donc sur les ordinateurs en eux-mêmes ainsi que sur leurs systèmes d'exploitation et nous irons ensuite voir du côté des logiciels.

- Ordinateur en tant que tel : cette fois encore, différentes informations sont laissées au constructeur. Nous pouvons par exemple citer les données de service lors de la création d'un compte. Comme expliqué précédemment, cela permettrait à l'entreprise de réaliser un lien entre toutes nos autres données et notre profil.

- Système d'exploitation de l'ordinateur : comme pour les smartphones, ces autres traces peuvent concerner toute l'utilisation que nous réalisons de l'ordinateur. Ainsi, nous laissons voir la quantité de batterie que nous utilisons, quand nous réalisons nos recharges, combien de temps l'appareil a été en veille, combien de temps et à quel moment il a été en utilisation, quels programmes ont été installés et utilisés, etc. Cette fois encore, il serait bien entendu très compliqué de lister la totalité des données que nous laissons tant celles-ci peuvent être nombreuses et complexes.

- Logiciels/programmes : ces outils peuvent récupérer une grande partie de nos traces numériques. Cette fois encore, nous parlons d'une utilisation plus personnelle de l'ordinateur. Nous n'allons également pas aborder les programmes basiques comme, par exemple, l'explorateur de fichier ou encore les logiciels de lecture de vidéos/musiques étant donné que ces programmes sont souvent intégrés dans le système d'exploitation. Voici donc d'autres logiciels qui récoltent nos données, cette sélection ayant également pour objectif d'illustrer le fait que de nombreuses traces sont laissées de façon générale au travers de ces éléments :

    - Environnements de développement intégré (Eclipse/PyCharm/BlueJ) : ces logiciels permettent de programmer dans différents langages informatiques. Lors du téléchargement de ceux-ci, il est parfois demandé d'introduire une adresse mail et quelquefois même d'autres informations. Il s'agit donc de données de service.

- Adobe Acrobat Reader : devenu un indispensable pour la manipulation de documents PDF, le logiciel peut se connecter à internet pour envoyer des informations quant à notre utilisation[11]. Cependant, il ne faut créer aucun compter ou fournir une quelconque information pour l'utiliser.

- Antivirus : maintenant sur beaucoup d'ordinateurs, la grande majorité de ces logiciels sont soit préinstallés, et donc dépendants du système d'exploitation, soit téléchargeables après inscription, de façon payante ou non. Cela implique donc des données de services ainsi que des informations plus larges sur l'ordinateur et l'utilisation qui en est réalisée.

- La suite Microsoft Office : même s'il existe un équivalent en logiciels libres, cette suite de logiciels est extrêmement populaire. Il est indiqué dans les paramètres de confidentialité que ces programmes analysent notre contenu.

- Logiciels de compression de données (WinRAR/7-Zip) : ces deux logiciels sont également devenus des incontournables, mais cette fois-ci pour la compression de fichiers et dossiers. Même si le premier est un logiciel propriétaire et le second est libre, ils fonctionnent tous les deux gratuitement (seulement au début pour WinRAR) et sans nécessiter d'information pour leur téléchargement. La seule particularité de WinRAR est qu'il dispose d'une version payante demandant une inscription.

- Plateformes de jeu : les grandes plateformes de jeux sur ordinateur permettent aux studios de développement de vendre leurs créations et offrent des facilités de gestion aux joueurs. En plus des données de service lors de l'inscription, nous donnons beaucoup d'autres informations, dont certaines de paiement. Au niveau des jeux, les entreprises savent ainsi quand nous jouons à chaque jeu, combien de temps nous y avons joué, où nous sommes arrivés, etc. Elles connaissent également la liste de contacts des utilisateurs, les jeux vus en magasins, ceux ajoutés à la liste des souhaits, ceux que l'utilisateur a décidé de suivre ou même ceux qu'il a consulté, etc. Avec cela, plusieurs entreprises arrivent à réaliser des systèmes de recommandation afin d'améliorer leurs services et réaliser plus de ventes. Certaines d'entre elles, comme Steam, récupèrent en effet une

[11] Adobe Acrobat Reader DC, 2019, « Politique de confidentialité », consulté le 30/052019, En ligne https://get.adobe.com/fr/reader/

commission lors de chaque transaction, en plus de demander un montant fixe au créateur lors de la publication de chaque jeu[12]. Il est également important de préciser que la situation est similaire pour les consoles de jeu.

Ainsi, les ordinateurs sont également la source de nombreuses traces numériques. Tout comme pour les smartphones, une inscription peut être demandée au niveau du constructeur et le système d'exploitation dispose également d'un grand pouvoir sur la récupération de nos traces. Concernant les logiciels, ceux-ci peuvent principalement récupérer nos données lors d'une éventuelle inscription mais cela peut bien entendu aller plus loin. A titre d'exemple, nous pouvons citer Steam pour qui il est capital de disposer d'un excellent système de recommandation. Tout comme c'était le cas pour l'application Play Store sur smartphone, la plateforme récupère une commission lors de chaque transaction et a donc tout intérêt à encourager celles-ci.

[12] Steam, 2018, « Community Market FAQ / Fees », consulté le 22/08/2018, En ligne https://support.steampowered.com/kb_article.php?ref=6088-UDXM-7214#steamfee

Navigation et services web

Dans cette partie, nous allons développer les traces spécifiques à la navigation et à certains services web. Dans les quatre premiers points, nous nous concentrerons sur des éléments essentiels pour se connecter sur le net, à savoir le fournisseur d'accès internet, l'adresse IP, le navigateur web et le moteur de recherche. Nous observerons ensuite quelques sites et services particuliers.

- Fournisseur d'accès internet : lors de chaque connexion sur un réseau, nous laissons des traces. S'il s'agit du notre, il y a souvent des données de services qui peuvent par ailleurs être combinées avec les données sur notre habitude de navigation internet. Si nous nous connectons sur un autre réseau, comme par exemple un wifi public, nous fournissons également des informations à cet autre endroit. Il est important de noter que beaucoup de réseaux gratuits ou ouverts demandent d'introduire une adresse mail, et parfois même d'autres informations, afin de pouvoir naviguer. Les réseaux wifi laissent également une trace supplémentaire sur nos appareils électroniques qui en gardent un historique. En effet, si le paramètre du wifi est activé, l'appareil cherche sans arrêt dans cette liste afin de voir s'il ne sait pas se connecter à un des réseaux qu'il capte.

- Adresse IP : il s'agit de la signature de tout appareil électronique connecté à internet. Celle-ci est attribuée par le fournisseur d'accès internet et est partagée à chaque site web sur lequel nous naviguons. N'importe quel site peut de cette manière reconnaitre l'appareil et même obtenir des informations supplémentaires comme la zone géographique dans laquelle nous nous trouvons. Cette dernière information peut notamment servir au site pour personnaliser un message ou une publicité. Nous savons cependant qu'il existe maintenant d'autres technologies plus avancées pour nous identifier (Grazia Cecere et al., 2015).

- Navigateur web : indispensable pour utiliser internet, il existe de nombreux navigateurs web différents exploitant plus ou moins nos données. Comme nous pouvons le voir dans la figure 2, le plus connu et le plus utilisé est Google Chrome, avec 60% des parts de marché.[13]

[13] StatCounter Global Stats, 2018, « Browser, OS, Search Engine including Mobile Usage Share », consulté le 12/08/2018, En ligne http://gs.statcounter.com/

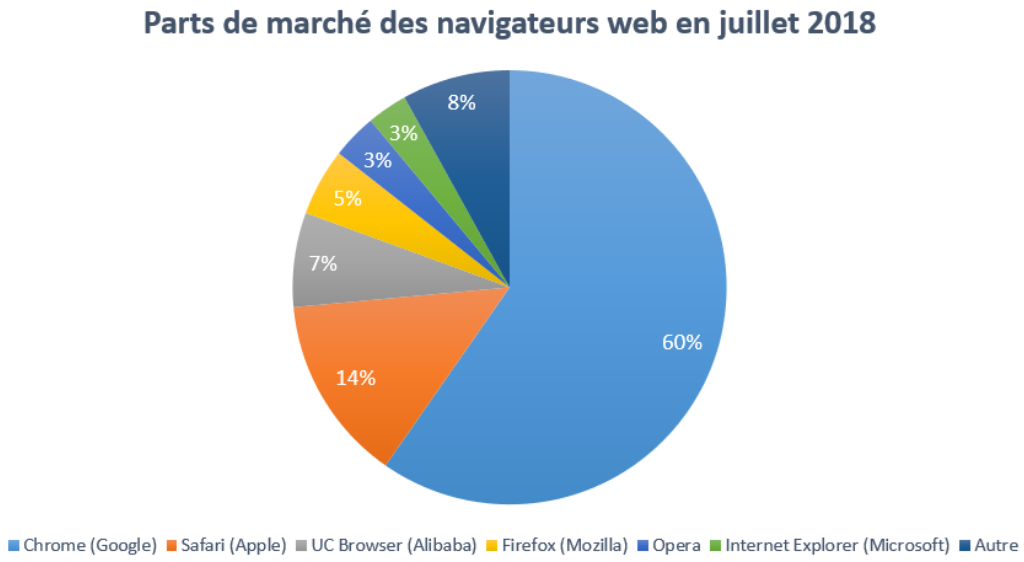

Figure 2 : Parts de marché des navigateurs web en juillet 2018

Les navigateurs peuvent tous obtenir des informations quant aux sites visités, au temps consacré à leur visualisation et aux personnalisations réalisées, comme par exemple l'ajout d'extensions. Ces dernières peuvent cependant servir d'outils afin de réduire son emprunte numérique. En effet, certaines extensions permettent notamment de bloquer les cookies utilisés par la plupart des sites. Elles aident ainsi à diminuer la quantité de données que l'on récolte sur nous. Cependant, il est maintenant très facile pour les sites web de nous identifier sans même utiliser de cookies et ce via le « *Device Finger Prints* », sujet abordé dans la revue de la littérature. En effet, chaque ordinateur dispose d'une signature, à savoir une série d'éléments qui lui permettent d'être reconnaissable dans une certaine mesure. En moyenne, seul 1 navigateur sur 3 millions dispose exactement de la même signature qu'un autre[14]. Il est donc maintenant totalement possible d'identifier quelqu'un sans même utiliser un compte ou des cookies, comme nous l'avons vu avec Grazia Cecere et al. (2015).

- Moteur de recherche : même s'il n'est pas obligatoire d'en utiliser un pour naviguer sur le web, son utilisation est devenue indispensable de nos jours. Cette fois encore, il existe de nombreux services exploitant plus ou moins nos données mais le plus important reste Google, avec plus de 90% des parts de marché[15] (cf. figure 3).

[14] Me and my Shadow (auteur anonyme), 2017, « Browser tracking », consulté le 12/08/2018, En ligne https://myshadow.org/browser-tracking

[15] StatCounter Global Stats, 2018, « Browser, OS, Search Engine including Mobile Usage Share », consulté le 12/08/2018, En ligne http://gs.statcounter.com/

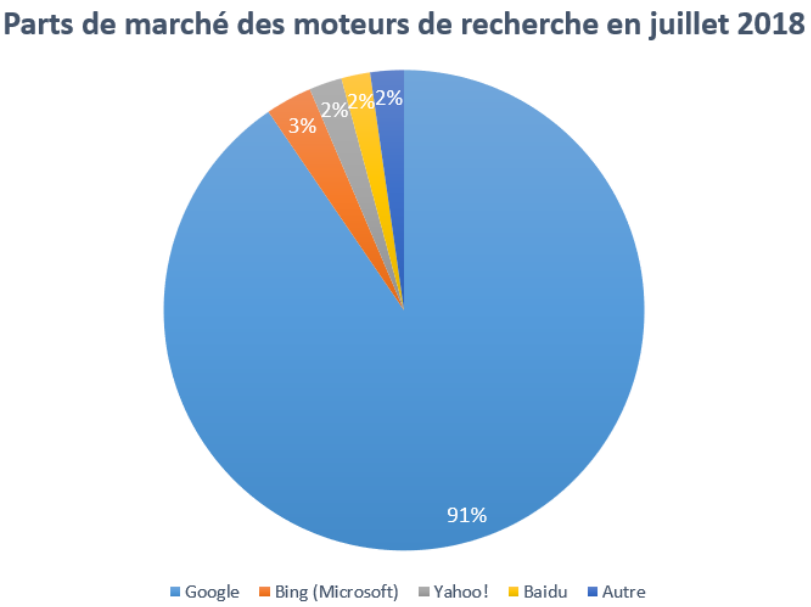

Figure 3 : Parts de marché des moteurs de recherche en juillet 2018

Un moteur de recherche a la possibilité de récolter beaucoup d'informations quant à nos habitudes de recherche, comme les mots-clés introduits ou encore les sites visités. Cette collecte fructueuse lui permet, par exemple, d'afficher de la publicité ciblée lors de la navigation ou encore d'être rétribué par certains sites pour accroitre leur visibilité.

Après avoir développé ces 4 éléments indispensables pour naviguer sur le web, nous pouvons examiner quelques sites et services particuliers :

- Services mails : il existe cette fois encore de nombreux fournisseurs de ce service avec des politiques de confidentialité différentes. Lors de la création de l'adresse mail, il est demandé de fournir diverses données de service. De plus, nous laissons beaucoup d'autres informations telles que les contacts, les messages envoyés, le fait qu'une adresse électronique personnelle envoie automatiquement ses mails à un autre adresse, etc. Il existe aussi différentes informations par rapport au nombre de messages envoyés, à qui ils sont envoyés, à quelle heure, aux objets, s'il y a des pièces jointes, etc. Le contenu du texte est également analysé comme par exemple lorsque nous mentionnons dans le contenu du texte « Pièce jointe » sans joindre quelque chose et qu'un message de confirmation mentionnant cet oubli apparait avant que le mail ne soit envoyé (cf. figure 4).

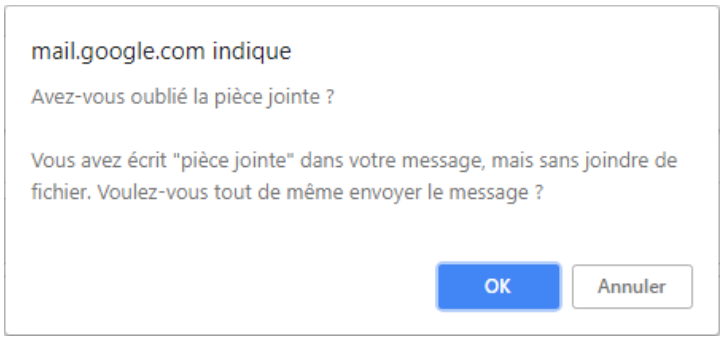

Figure 4 : Fenêtre pop-up du service de messagerie électronique Gmail

- Sites d'hébergement de vidéos : à l'heure actuelle, il existe d'innombrables sites d'hébergement de vidéos, le plus populaire étant YouTube, site racheté par Google en 2006[16]. De plus, nous pouvons même inclure des plateformes aux concepts plus larges telles que edX, le site de cours en ligne, même s'il est incorrect de résumer ce service à du simple hébergement de vidéos. Dans de tels sites, en plus des données de services fournies lors de la création d'un compte, nous divulguons beaucoup d'informations quant à notre consommation de vidéos. Nous pourrions ainsi citer des données quant aux vidéos regardées, au temps consacré, aux chaines que nous suivons, à notre intérêt pour les publicités en décidant de les regarder ou non, à d'éventuels commentaires et évaluations laissés, etc. Tous ces éléments donnent des informations cruciales qui servent à compléter notre profil. Grace à cela, il est possible pour les hébergeurs de cibler de la publicité, une importante source de revenus[17]. De plus, notre profil leur permet également de nous inciter à regarder des vidéos, en utilisant notamment un système de recommandation, afin de pouvoir exposer encore plus de publicités ciblées.

- Commerce électronique : secteur en plein essor, le commerce électronique est devenu un incontournable. Les deux principaux sites sont Alibaba et Amazon, deux géants respectivement chinois et américain, mais cette catégorie comprend également les chaines de commerce plus traditionnelles proposant une boutique en ligne. Cette fois encore, il est nécessaire de fournir des informations de services, d'autant plus que ces produits sont censés être livrés à une adresse. Par la suite, d'importantes quantités de

[16] Hickey M., 2006, « Google Buys YouTube for $1.65 Billion », consulté le 12/08/2018, TechCrunch, En ligne https://techcrunch.com/2006/10/09/google-buys-youtube-for-165-billion/

[17] Dutta A., 2019, « YouTube Business Model | How Does YouTube Make Money? », consulté le 30/05/2019, Feedough, En ligne https://www.feedough.com/youtube-business-model-how-does-youtube-make-money/

données sont récoltées. Nous pouvons bien entendu citer les types de produits achetés, à quel moment, seuls ou groupés, etc. mais cela serait réducteur. En effet, ces sites retiennent également les mots-clés introduits dans la barre de recherche, les produits sur lesquels nous avons cliqué ou non, combien de temps nous sommes restés sur la page, si nous les avons ajoutés dans notre panier ou dans une liste particulière, etc. De cette manière, ces commerces arrivent à construire des systèmes de recommandation hautement performants. Grâce à ces méthodes, ils maximisent leurs chances de réaliser des ventes et augmentent le potentiel de sérendipité, à savoir le fait de faire découvrir des produits auxquels les utilisateurs n'auraient pas pensé. Selon Viktor Mayer-Schönberger (2014), environ un tiers des revenus de l'entreprise Amazon serait d'ailleurs généré par leur système de recommandation. Cela ne se limite bien entendu pas aux plus grands sites étant donné que, selon une étude de *Econsultancy and Monetate* (2013), 94% des sites de e-commerce considèrent les systèmes de recommandation comme un avantage compétitif critique à implémenter.

- Service de streaming vidéo : avec plus de 125 millions d'abonnés[18], Netflix est le plus grand site de streaming au monde. Tout comme pour le commerce électronique, les systèmes de recommandation ont une place très importante. En effet, le site n'utilise pas de publicité mais compte sur les abonnements de ses utilisateurs. Il a donc tout intérêt à leur fournir des propositions pertinentes pour les inciter à continuer à payer leur abonnement. Les traces récoltées sont bien entendu des données de service et d'innombrables informations quant à l'utilisation du site. Ainsi, chaque fait et geste est analysé, notamment les recherches, les liens sur lesquels nous avons cliqué ou non, le temps passé à regarder une série ou un film, si nous avons bel et bien regardé cela jusqu'à la fin, notre éventuelle évaluation, etc.

- Service de streaming musical : tout comme son équivalent vidéo, le streaming musical est un marché très important et en plein essor. Comme principaux acteurs, nous pouvons citer Spotify, Deezer, Apple Music ou encore Google Music. Tous fonctionnent avec un système d'abonnement et une éventuelle publicité. Ainsi, il est particulièrement important pour ces sites d'avoir un bon système de recommandation afin d'inciter les utilisateurs à souscrire un abonnement. A titre d'exemple, nous pouvons citer Spotify

[18] Lawler R., 2018, « Netflix subscriber count hits 125 million », consulté le 12/08/2018, engadget, En ligne https://www.engadget.com/2018/04/16/netflix-subscriber-count-hits-125-million/?guccounter=1

Discover Weekly qui génère une playlist personnalisée chaque semaine et qui est particulièrement réputé pour sa qualité[19]. Les données récoltées sont, quant à elles, classiques, à savoir certaines générées lors de la création du compte et d'autres pendant l'utilisation du service. Pour ces dernières, nous pouvons citer en exemple les musiques écoutées, celles enregistrées dans telle ou telle playlist, si nous les avons écoutées totalement, les personnes que nous suivons, etc.

- Service de cartographie en ligne : le site le plus connu est Google Maps, d'autant plus qu'il sert de GPS sur deux tiers des smartphones[20]. Les traces laissées concernent principalement la position géographique et les endroits visités, éléments similaires dans l'utilisation de certains boitiers GPS. Il est cependant important de noter que Google a une place très importante dans la plupart des services que nous avons vu dans cette catégorie. Cela devient ainsi particulièrement intéressant pour le géant américain si les utilisateurs multiplient ses services car cela lui permet de recouper les données. Il crée de cette manière de nouvelles informations ou du moins des informations plus complètes, comme nous l'avons vu avec Bruce Schneier (2015) et les données dérivées.
- Services de commande de repas : en se servant de tels services, nous devons donner des informations quant à notre position, aux endroits où nous aimons manger, à la nourriture appréciée, etc. Il est également intéressant de souligner que ces applications gagnent de l'argent en récupérant une commission lors de chaque commande. Il est donc rassurant de se dire qu'elles ne sont pas forcées d'utiliser nos données personnelles pour subsister. Elles se permettent cependant d'envoyer des offres personnalisées et de réaliser de la publicité ciblée.

[19] Pasick A., 2015, « The magic that makes Spotify's Discover Weekly playlists so damn good", consulté le 25/05/2019, Quartz, En ligne https://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/

[20] Panko R., 2018, « The Popularity of Google Maps: Trends in Navigation Apps in 2018 », consulté le 30/05/2019, The Manifest, En ligne https://themanifest.com/app-development/popularity-google-maps-trends-navigation-apps-2018

Ainsi, les smartphones et ordinateurs sont bel et bien de grands pourvoyeurs de traces numériques, notamment au travers de la navigation et de certains services web. Nous sommes en effet facilement identifiables avec notre adresse IP ou encore notre navigateur et nous laissons d'importantes traces à notre fournisseur d'accès internet ainsi qu'à notre moteur de recherche. Lors de l'utilisation d'un produit nécessitant un compte, nous laissons différentes données de services mais également d'utilisation du site en question. Nous pouvons par exemple citer les services mail, les sites d'hébergement de vidéos, le commerce électronique ainsi que les services de streaming vidéo et musical. Enfin, il ne faut également pas négliger les services de cartographie en ligne qui peuvent profiter de la localisation GPS et dégager d'importantes données de comportement, ou encore les services de commande de repas.

Réseaux sociaux

Même s'ils auraient pu être classés dans la catégorie précédente, les réseaux sociaux bénéficient de leur propre identité tant ceux-ci fonctionnent différemment (Grazia Cecere et al., 2015) et jouent un rôle important dans les traces numériques que nous laissons. Afin de choisir lesquels analyser, nous pouvons observer dans la figure 5 les réseaux sociaux les plus utilisés[21] :

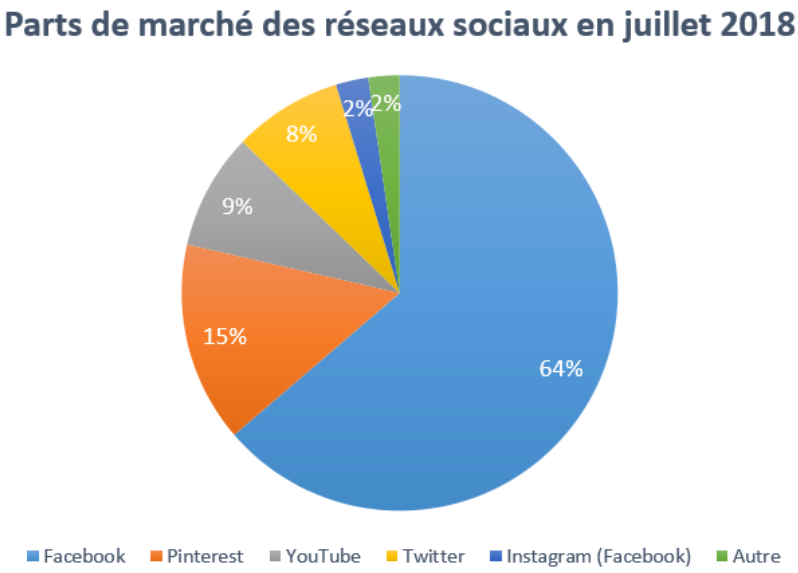

Figure 5 : Parts de marché des réseaux sociaux en juillet 2018

Comme nous pouvons le voir, Facebook détient la majorité des parts de marché, d'autant plus qu'il possède également le réseau social Instagram. Il est également intéressant de noter que YouTube est ici considéré comme un réseau social mais nous n'en parlerons plus car nous l'avons déjà abordé dans la catégorie précédente. La présence de ce site offre tout de même la possibilité de comparer son importance à celle d'autres acteurs comme Facebook.

La particularité des réseaux sociaux est qu'ils font largement usage des données confiées, donc divulguées volontairement sur un site que nous ne possédons pas, et des données annexes, divulguées par des tierces personnes. Nous examinerons ainsi Facebook, Twitter et finalement les réseaux centrés sur le partage de photos, à savoir Pinterest, Instagram et Snapchat.

[21] StatCounter Global Stats, 2018, « Social Media Stats Worldwide », consulté le 22/08/2018, En ligne http://gs.statcounter.com/social-media-stats

- Facebook : l'entreprise connait les contacts d'un individu, les pages aimées, le nombre de statuts, de commentaires, de mentions « j'aime », les événements auxquels il participe, le nombre et les types de groupes dans lesquels il se trouve ainsi que son activité dans ceux-ci, des informations sur d'éventuelles pages dont il serait administrateur, ses informations personnelles volontairement données afin de compléter son profil, etc. Cependant, il faut également compter toutes les données annexes, à savoir les publications où il a été mentionné, les photos et statuts dans lesquels il a été identifié, ou encore toutes les données dérivées obtenues par exemple grâce à l'analyse de réseau.

  Pour avoir une vision d'ensemble sur un acteur aussi important que Facebook, nous pouvons utiliser notre droit d'accès à une copie de l'ensemble des données nous concernant que possède l'entreprise. Un fichier de 2,5 Go contenant les données d'utilisation du service pendant 6 ans est ainsi téléchargé. Dans les éléments sortant de la liste dressée ci-dessus, il est ainsi intéressant de savoir que l'entreprise a accès à l'ensemble des contacts du smartphone, aux appels et aux SMS. Elle garde en mémoire d'importantes données de publicité, comme par exemple la liste des annonceurs lorsque nous avons cliqué sur les publicités, l'ensemble des demandes d'amitié envoyées, reçues, refusées, etc. et une liste complète des moments et endroits où nous nous sommes connectés, avec quels appareils et adresses IP. L'explication de l'ensemble du fichier est disponible en annexe 2. Toutes ces informations permettent de cibler de la publicité, principale source de revenus du géant américain (98% en 2018[22]).

- Twitter : même s'il est moins populaire que Facebook, ce réseau créé en 2006 a réussi à obtenir une place de référence dans ce marché. Tout comme Facebook, différentes données sont fournies lors de la création du profil. Par après, beaucoup d'autres éléments sont enregistrés comme, par exemple, les personnes suivies, celles qui nous suivent, les retweets, le nombre de tweets, quand ceux-ci sont réalisés, qui y est mentionné, avec quels hashtags, le nombre de caractères et la nature des mots utilisés, etc. Il est également intéressant d'observer qu'il est impossible de supprimer instantanément un compte de façon définitive, particularité commune à beaucoup de réseaux sociaux. Il faut en effet désactiver le compte qui reste potentiellement

[22] Facebook, 2019, « Annual Report 2018 », consulté le 26/05/2019, En ligne https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/2018-Annual-Report.pdf

réactivable durant un mois. Le compte est finalement définitivement supprimé après un mois s'il n'y a pas de tentative de connexion.

- Réseaux sociaux centrés sur le partage de photos : ce dernier point comporte tous les réseaux sociaux qui mettent au centre de leur concept le partage de photos, à savoir Pinterest, Instagram et Snapchat. Nous nous permettons ainsi d'ajouter Snapchat dans cette catégorie même s'il n'était pas présent dans la figure 5 étant donné qu'il reste très populaire et dispose de caractéristiques similaires. Ces réseaux ont la particularité d'être plus centrés sur les smartphones étant donné leur concept : la prise de photos rapide et facile à partager. Ils récupèrent donc les informations quant aux photos envoyées, à qui elles sont destinées et à quelle heure, aux éventuels filtres utilisés, aux personnes les ayant consultées, aux amis, etc. Comme pour la plupart des réseaux sociaux, nous fournissons des données de service, confiées et annexes. Il est également important de noter que, lors de l'installation sur smartphone, Instagram et Snapchat demandent l'accès à la position GPS, aux contacts, aux SMS et aux appels, soit des fonctionnalités que l'on peut difficilement justifier pour de telles applications. Quant à Pinterest, seuls les accès aux contacts et à la position GPS sont demandés, bien entendu en plus des autorisations classiques, afin de pouvoir fournir leur service, comme par exemple l'accès à l'appareil photo.

Ainsi, les réseaux sociaux observent nos moindres faits et gestes réalisés sur leur plateforme afin de s'en servir notamment pour de la publicité ciblée. De plus, la majorité de ceux-ci demandent des autorisations d'accès sur smartphones et récupèrent de la sorte les contacts, SMS et appels. On peut réaliser un parallèle avec les sites de rencontre. Ceux-ci récupèrent en effet les données fournies par leurs utilisateurs afin d'afficher de la publicité ciblée ou d'améliorer leurs services.

Cadre hors web

Dans cette dernière catégorie, nous allons découvrir des éléments hors cadre direct du web. Nous verrons tout d'abord le domaine bancaire, notamment via nos traces laissées par les cartes de paiement, le commerce dans sa version plus traditionnelle et les nombreux moments où il nous est demandé de donner une information d'identité.

- Informations bancaires : ce point est particulier car les informations bancaires sont censées être confidentielles. Nous allons cependant l'aborder brièvement tout en gardant cela à l'esprit. Nous pouvons considérer qu'il y a des données de services liées aux informations fournies lors de l'ouverture d'un compte mais aussi lors d'un paiement en ligne. De plus, nous fournissons diverses données à notre banque et aux commerces où nous réalisons des achats lors de chaque payement effectué avec une carte bancaire.

- Cartes de fidélité : une grande partie des magasins proposent désormais une carte de fidélité à leurs clients. Cet outil permet de réaliser un suivi de ceux-ci et de leurs habitudes d'achats, tout comme le ferait un commerce en ligne comme Amazon ou Alibaba. Nous pourrions même étendre ce point à tous les formulaires papier que nous remplissons, par exemple dans le cas d'une enquête de satisfaction qui n'est pas anonyme.

  Il est donc très important pour ces établissements d'observer des corrélations et de réaliser des catégories dans lesquelles classer leurs clients. Pour illustrer cela, nous pouvons utiliser l'exemple suivant : une grande chaine de supermarchés aux Etats-Unis a observé que les personnes achetant des bières achètent aussi des couches culottes. Même si cet exemple a été largement repris et modifié au cours de l'histoire, l'explication la plus consensuelle est que les jeunes parents voient souvent la mère rester à la maison et envoient donc le père faire les courses. Celui-ci achète ainsi les produits de première nécessité selon lui, à savoir des couches et de la bière pour le match qu'il va regarder à la télévision le soir. La stratégie serait donc de modifier le positionnement dans le magasin pour inciter ces personnes à acheter d'autres produits, comme par exemple des cacahuètes et des chips.

- Autres services : il existe de nombreux services pour lesquels nous laissons certaines informations et ce sans passer par le web. Nous pouvons par exemple citer l'achat d'une place nominative de spectacle, la réservation d'un billet d'avion en agence, l'achat d'un titre de transport nominatif, un abonnement à un magazine, etc. Cela concerne ainsi chaque moment où il nous est demandé de fournir un élément correspondant à notre identité.

Comme nous l'avons vu, nous laissons certaines traces et signatures avec nos informations bancaires, même s'il faut garder à l'esprit que celles-ci sont confidentielles. De plus, tout comme sa version en ligne, le commerce traditionnel récupère nos données lorsque nous utilisons des cartes de fidélité afin d'augmenter ses ventes. Nous laissons également des données à chaque fois qu'une information d'identité nous est demandée pour profiter d'un service.

En conclusion, nous avons abordé les traces spécifiques aux smartphones, à savoir celles propres à l'appareil en tant que tels et à son système d'exploitation. Nous avons vu le rôle de l'opérateur téléphonique, de la position GPS et nous nous sommes concentrés sur certaines applications, à savoir Play Store, Camera, Agenda, National, Shazam et MyProximus.

Par après, nous avons décrit les traces spécifiques aux ordinateurs. Cette fois encore, nous avons vu celles propres à l'ordinateur en tant que tel et celles liées à son système d'exploitation. Nous avons abordé différents programmes, à savoir les environnements de développement intégré, Adobe Acrobat Reader, les antivirus, la suite Office, des logiciels de compression de données et les plateformes de jeu.

Nous avons ensuite cherché les traces communes aux smartphones et aux ordinateurs, soit la navigation et les services web. Nous avons ainsi pu voir l'importance du fournisseur d'accès internet et de l'adresse IP qu'il nous attribue. Pour clôturer cette première partie sur les éléments indispensables à la navigation, nous sommes passés aux navigateurs web et aux moteurs de recherche. Finalement, nous avons abordé quelques sites et services souvent utilisés, à savoir les services de messagerie électronique, les sites d'hébergement de vidéos, le commerce électronique, les sites de streaming vidéo et musical, les services de cartographie en ligne et enfin les services de commande de repas.

Par la suite, nous avons pu voir d'autres éléments également communs aux smartphones et aux ordinateurs mais qui méritaient leur propre catégorie, à savoir les réseaux sociaux. Nous avons pu les séparer en trois grands acteurs, à savoir Facebook, possédant la majorité des parts de marché, Twitter, et enfin les réseaux centrés sur le partage de photos comme Pinterest, Instagram et Snapchat.

Finalement, nous avons abordé une dernière catégorie sortant du cadre du web. Nous avons ainsi pu voir nos traces numériques au travers de nos informations bancaires, même si celles-ci sont confidentielles. Nous avons également réalisé une partie sur les cartes de fidélité et sur toutes les données que nous divulguons pour profiter de certains services.

Techniques, habitudes et outils à mettre en place pour limiter ses traces

Après avoir listé les traces numériques que nous laissons, nous allons maintenant voir comment les supprimer, ou du moins les limiter. Comme nous l'avons vu précédemment, il existe plusieurs types de données et donc de traces. Nous avons ainsi déjà pu observer qu'il est possible d'en limiter certaines mais que d'autres sont beaucoup plus complexes à supprimer, comme par exemple les données de service et les données annexes. Eviter les premières nous obligerait à passer par de fausses informations, technique présentant ses limites, ou par le simple arrêt de l'utilisation du service en question. Les données annexes sont quant-à-elles totalement indépendantes de notre contrôle et présentent également des difficultés. Nous développerons ainsi différents outils mais également des techniques et habitudes, une partie des problèmes n'étant pas solutionnable uniquement via des outils.

Dans ce chapitre, nous allons donc reprendre la structure de notre chapitre précèdent afin d'exposer les solutions pour limiter nos traces. Nous parlerons tout d'abord des smartphones et de leurs difficultés au niveau des systèmes d'exploitation. Nous passerons ensuite aux ordinateurs où nous développerons diverses solutions et nous compléterons avec la navigation et les services web. Nous verrons également le cas des réseaux sociaux, des informations bancaires et des cartes de fidélité. Nous aborderons finalement la suppression de nos traces actuelles, afin de limiter la création de données dérivées.

Smartphones

Dans cette partie, nous essayerons de limiter les traces numériques spécifiques aux smartphones, à savoir au niveau de l'appareil en tant que tel, son système d'exploitation, l'opérateur téléphonique, la localisation GPS et les applications.

- Smartphone en tant que tel : comme nous l'avons vu, les traces laissées concernent simplement l'éventuelle création d'un compte afin de bénéficier de certains avantages ou profiter de différents services. Les solutions sont donc assez simples, à savoir ne pas créer ce compte ou acheter un smartphone conçu spécialement pour préserver la vie privée. Plus de détails sur un tel appareil sont donnés au point suivant.

- Système d'exploitation : il s'agit sans doute de la plus grande difficulté des smartphones. Il n'existe en effet que 4 grands systèmes d'exploitation (OS) et tous appartiennent à des entreprises majeures : Android (Google), iOS (Apple), Windows 10 Mobile et Blackberry OS. Ceux-ci sont d'ailleurs loin d'être répartis équitablement étant donné que Android possède, en juillet 2018, 77% des parts de marché et iOS 19%[23]. Il existe cependant quelques alternatives que nous allons développer, même si celles-ci peuvent toujours récupérer certaines données. Nous pouvons par ailleurs garder à l'esprit que cela sera toujours dans une moindre mesure que les 4 grands systèmes d'exploitation. Afin de ne pas présenter la totalité des projets, voici une sélection en fonction de leur pertinence et de leur popularité :

    - Firefox OS : l'entreprise Mozilla annonce en 2012 qu'elle développe un système d'exploitation libre et open source sur smartphone. Le projet est cependant abandonné en 2016, son arrivée étant trop tardive sur un marché déjà bien établi et largement dominé par Google et Apple. Le code source du système est par ailleurs toujours disponible en téléchargement.

[23] StatCounter Global Stats, 2018, « Mobile Operating System Market Share Worldwide », consulté le 15/08/2018, En ligne http://gs.statcounter.com/os-market-share/mobile/worldwide

- Ubuntu Touch : il s'agit d'un système d'exploitation annoncé début 2013 par l'entreprise Canonical[24]. Son développement est cependant arrêté en avril 2017 mais le projet a été repris par la communauté. Il est pour le moment loin d'être abouti et il n'est supporté que par quelques rares appareils. De plus, le passage d'un OS vers Ubuntu Touch signifie également la perte de toutes les données.

- LineageOS : ce système d'exploitation open source, toujours en développement, est annoncé en décembre 2016. Il fait suite à un précédent projet, le CyanogenMod. Il a l'avantage d'être supporté par de très nombreux appareils, même si ceux-ci doivent forcément avoir été développés pour Android. Son inconvénient principal est qu'il est très compliqué de changer le système d'exploitation de son smartphone pour le remplacer par celui-ci. Cette difficulté n'est par ailleurs pas propre à cet OS mais est commune à l'ensemble des systèmes alternatifs.

- Librem 5 : il s'agit d'un smartphone développé par l'entreprise Purism et équipé d'un système d'exploitation open source. Sa conception est spécialement prévue pour garantir la sécurité et l'intégrité des données personnelles de son propriétaire. Afin de mener à bien son projet, l'entreprise réalise une importante campagne de financement participatif et réussit largement à dépasser son objectif initial s'élevant à 1,5 millions de dollars. Le point faible de ce smartphone est son prix, à savoir 649$[25].

- Opérateurs téléphoniques : il s'agit également d'une importante difficulté pour les smartphones. Ces acteurs peuvent récolter beaucoup d'informations qui sont ensuite liées à notre profil. La première étape pour diminuer ses traces numériques est de ne pas se créer de compte chez son opérateur, même s'il deviendrait impossible de profiter des avantages qu'il peut proposer. Voici ensuite d'autres solutions qui pourraient être mises en place :

[24] Le Monde Informatique (auteur anonyme), 2013, « Canonical lancera Ubuntu Touch le 17 octobre », consulté le 15/08/2018, En ligne https://www.lemondeinformatique.fr/actualites/lire-canonical-lancera-ubuntu-touch-le-17-octobre-55098.html

[25] Purism, 2019, « Librem 5 », consulté le 30/05/2019, En ligne https://shop.puri.sm/shop/librem-5/

- Carte SIM prépayée : avec une carte prépayée, une grande partie des traces est ainsi supprimée. Il ne faut plus se créer de compte mais il est toujours nécessaire de présenter une pièce d'identité lors de l'achat de cette carte. Il s'agit en effet d'une loi votée en 2016 afin de « mettre fin à l'anonymat des cartes SIM prépayées »[26].

- Réseaux pirates : il est maintenant possible de créer des réseaux téléphoniques pirates afin de ne plus passer par des opérateurs classiques. Une telle solution est cependant compliquée à mettre en place et est illégale. Il existe par ailleurs depuis quelques années des boitiers complets contenant tout le matériel nécessaire à cela[27]. Un tel réseau est cependant plus utilisé par des *hacktivists* que par des personnes voulant réduire leur emprunte numérique.

- Arrêt total de l'utilisation : cette dernière solution est extrême et signifierait se priver d'un quelconque téléphone. Elle peut cependant être considérée comme acceptable si notre utilisation du smartphone se limite à l'usage de ses autres fonctionnalités. Il est en effet totalement possible pour quelqu'un d'envoyer des messages et passer des appels avec une connexion internet, en se privant bien évidemment de l'internet mobile. Cependant, ayant réalisé une enquête afin de clarifier plusieurs aspects de ce mémoire, il ressort que seul un répondant sur 40 n'utilise jamais les SMS (1 répondant sur 20 pour les appels).

- Localisation GPS : cet élément est dépendant du système d'exploitation du smartphone et de ses applications. Ainsi, si un autre OS est utilisé et que les services dont nous nous servons ne récupèrent pas ces données, le problème est résolu. Il faut cependant bien faire attention à employer les bonnes applications.

[26] Tamigniau M., 2016, « La fin des cartes SIM anonymes, prisées des malfrats et des terroristes, est imminente… mais elle pourrait être inutile », consulté le 26/08/2018, RTL Info, En ligne https://www.rtl.be/info/vous/temoignages/la-fin-des-cartes-sim-anonymes-prisees-des-malfrats-et-des-terroristes-est-imminente-mais-elle-pourrait-etre-inutile-821173.aspx

[27] Lapprand M., 2014 « Piratebox, ou comment créer son propre réseau Internet », consulté le 16/08/2018, Réseau International, En ligne https://reseauinternational.net/piratebox-ou-comment-creer-son-propre-reseau-internet/

- Applications : en cas de changement de système d'exploitation, toutes les applications spécifiques à son utilisation sont déjà remplacées. Pour les autres, il est très probable de trouver leur équivalent en version open source et respectueuse des données personnelles. Il existe ainsi une application remplaçant le catalogue Play Store, à savoir F-Droid[28]. Celle-ci propose une sélection d'applications toutes open source comme, par exemple, DuckDuckGo, un navigateur web et moteur de recherche conçu pour ne récupérer aucune donnée personnelle et pour bloquer les traqueurs publicitaires.

- Utiliser un ancien téléphone : si nous nous rendons compte que nous n'exploitons que très peu les fonctionnalités d'un smartphone, il pourrait alors être pertinent de revenir à un ancien modèle de téléphone beaucoup plus basique. Cette option ne résout par ailleurs pas le problème de l'opérateur téléphonique mais permettrait de grandement réduire ses traces.

Ainsi, il faut éviter au maximum de se créer des comptes auprès du fabricant du smartphone ou de l'opérateur téléphonique. Concernant les systèmes d'exploitation, il existe quelques alternatives libres et open sources mais celles-ci sont souvent compliquées à installer. Afin de pallier ce problème, un smartphone a spécifiquement été créé pour protéger les données de ses utilisateurs. L'opérateur téléphonique est quant à lui difficile à contourner, contrairement aux applications qui présentent souvent des équivalents plus respectueux des données personnelles.

[28] Lada J., 2014, « Un système d'exploitation 100% open source est-il possible ? », consulté le 14/08/2018, PhonAndroid, En ligne http://www.phonandroid.com/systeme-exploitation-entierement-open-source-possible.html

Ordinateurs

Dans cette partie, nous allons décrire des solutions qui permettent de réduire les traces numériques spécifiques aux ordinateurs. Nous verrons ainsi le problème de l'ordinateur en tant que tel et de son système d'exploitation. Nous finirons en abordant le sujet des logiciels installés.

- Ordinateur en tant que tel et système d'exploitation : concernant l'appareil, il faudrait simplement ne pas créer de compte client auprès du fabriquant. De cette manière, il serait impossible de lier notre profil aux données que nous laissons. Son système d'exploitation est, quant à lui, plus facile à solutionner que celui du smartphone. Il existe de nombreuses alternatives aux géants de l'informatique, plus faciles à installer que leur équivalent sur téléphone mobile. Tout comme pour les smartphones, le marché des systèmes d'exploitation est largement dominé par quelques acteurs dont Windows (Microsoft) qui possède 83% des parts de marché en juillet 2018 et OS X (Apple), avec 13%[29]. Voici une sélection de systèmes libres initialement réalisée par le journal Le Monde[30] :
    - Ubuntu : sorti en 2004, il s'agit de la version PC de Ubuntu Touch, présentée précédemment. Comme la majorité des systèmes d'exploitation, beaucoup de logiciels sont installés par défaut, dont un équivalent de Microsoft Office. Il sert de base à de nombreux autres systèmes et des mises à jour sont toujours réalisées régulièrement.
    - Mint : il s'agit d'un OS basé sur Ubuntu sorti en 2006. Sa particularité réside dans son interface simple et la différence majeure avec Ubuntu est l'utilisation de logiciels propriétaires. Le système est d'ailleurs absent de la liste des distributions soutenues par la Free Software Foundation[31], une

[29] StatCounter Global Stats, 2018, « Desktop Operating System Market Share Worldwide », consulté le 16/08/2018, En ligne http://gs.statcounter.com/os-market-share/desktop/worldwide
[30] Leloup D., 2015, « 5 systèmes d'exploitation pour snober Windows 10 (et Mac OS) », consulté le 16/08/2018, Le Monde, En ligne https://www.lemonde.fr/pixels/article/2015/08/04/cinq-systemes-d-exploitation-pour-snober-windows-10-et-mac-os_4710726_4408996.html
[31] GNU Operating System - Free Software Foundation, 2018, « Explaining Why We Don't Endorse Other Systems », consulté le 16/08/2018, En ligne https://www.gnu.org/distros/common-distros.html

organisation américaine créée en 1985 et dont le but est de promouvoir les logiciels libres ainsi que la défense des utilisateurs.

- ArchLinux : créé en 2002, ce système s'adresse à des utilisateurs aux connaissances plus pointues en informatique. Son utilisation est en effet assez complexe mais permet une customisation presque totale. Pour illustrer cela, le système ne comporte par exemple aucun environnement graphique afin de ne pas masquer son fonctionnement. Pour les mêmes raisons que Mint, ArchLinux ne figure pas sur la liste des distributions soutenues par la Free Software Foundation.

- Elementary : ce système d'exploitation sorti en 2011 est lui aussi dérivé de Ubuntu. Son interface est simple et légère, permettant de fonctionner sur des modèles d'ordinateurs moins puissants. Sa particularité est qu'il utilise une interface graphique très proche de OS X, le système d'exploitation des ordinateurs Apple.

- Tails : il s'agit du système qui nous intéresse particulièrement dans le cadre de ce mémoire. Il a été conçu pour protéger la vie privée et a notamment été utilisé par Edward Snowden pour communiquer avec les journalistes. Son nom provient de l'acronyme « The Amnesic Incognito Live System » et sa particularité est qu'il s'installe sur un support amovible, comme une clé USB. Il suffit ainsi de démarrer un ordinateur avec cette clé connectée afin de mettre en route le système. Il ne laisse pas de traces sur l'appareil, il ne conserve aucune donnée et, en cas d'urgence, il suffit de retirer la clé pour l'arrêter. Pour communiquer, il passe par le réseau Tor, un système décentralisé mondialement reconnu pour son anonymat[32], et chiffre les fichiers, les mails et les messages.

  Comme les autres OS, il dispose de logiciels préinstallés comme un programme de retouche d'images et une suite bureautique. L'outil est déjà largement utilisé et est même recommandé par Reporters sans frontières ou encore la Freedom of the Press Foundation. La NSA, quant à elle, qualifie

[32] Heffernan V., 2010, « Granting Anonymity », consulté le 26/05/2019, New York Times, En ligne https://www.nytimes.com/2010/12/19/magazine/19FOB-Medium-t.html?searchResultPosition=2

l'outil de « dispositif de sécurité des communications préconisé par des extrémistes, sur des forums extrémistes »[33].

- Logiciels/Programmes : comme pour les smartphones, il est possible de réduire ses traces numériques en préconisant des logiciels libres et open source. Certains ne demandant que quelques informations, comme par exemple une adresse mail, les traces peuvent être évitées via de faux renseignements. Pour d'autres logiciels, dont certains payants, il est sans doute possible de trouver leur équivalent en version libre. Pour aider dans cette tâche, Framasoft[34] voit le jour en novembre 2001. Son objectif s'articule autour de trois axes principaux : la promotion, la diffusion et le développement de logiciels libres. Le site propose non seulement des logiciels mais également de la culture et des services libres. Parmi les plus notables, nous pouvons citer une importante base de données gratuite de livres, de vidéos et de musique, des logiciels de traitement de texte, d'édition d'image, d'agendas, de discussion vidéo, d'actualité ou encore un service d'hébergement de fichiers en ligne. Il ne s'agit que d'un échantillon de la liste d'alternatives disponibles, catalogue qui continue de s'agrandir avec le temps. Nous allons maintenant réutiliser la liste réalisée dans le chapitre précédent afin de s'assurer que chaque logiciel soit bien couvert :
    - Environnements de développement intégré : ils ne demandent pas toujours des informations lors du téléchargement. Si c'est le cas, il suffit d'en donner des fausses.
    - Adobe Acrobat Reader : il ne demande pas non plus d'information.
    - Antivirus : la plupart des antivirus ont une politique très stricte en termes d'utilisation des données privées. De plus, certains outils comme Tails ne nécessitent pas l'utilisation d'un autre système pour se protéger lors de la navigation sur le web.

[33] Guiton A., 2014, « Tails, l'outil détesté par la NSA », consulté le 16/08/2018, Le Monde, En ligne https://www.lemonde.fr/pixels/article/2014/11/20/tails-l-outil-deteste-par-la-nsa-qui-veut-democratiser-l-anonymat-en-ligne_4514650_4408996.html

[34] Framasoft, 2018, « Page portail du réseau », consulté le 16/08/2018, En ligne https://framasoft.org/fr/

- La suite Office : la solution la plus populaire pour remplacer cette suite de logiciels est d'utiliser OpenOffice.org.[35], une suite bureautique libre et gratuite. Elle est composée de 6 programmes différents, à savoir un logiciel de traitement de texte, un tableur, un logiciel de présentation sous forme de diapositives, un programme de création de bases de données, un outil de dessin vectoriel et un outil de création de formules mathématiques. De plus, certains systèmes d'exploitation alternatifs comme Tails ont également une suite bureautique déjà installée.

- Logiciels de compression de données : tout comme pour les environnements de développement intégré, ils ne demandent pas toujours des informations lors du téléchargement et, si c'est le cas, il suffit d'en donner de fausses.

- Plateformes de jeu : même s'il existe de nombreux sites alternatifs, ceux-ci ne garantissent absolument pas une absence d'utilisation de nos données personnelles. Nous avons ainsi différentes solutions plus ou moins efficaces :

    - Trouver une alternative avec une politique de confidentialité plus stricte. Cette solution semble être la moins efficace. De plus, il est peu probable de trouver une telle alternative disposant d'un catalogue aussi important que, par exemple, Steam. Il existe cependant de nombreux sites répertoriant des jeux gratuits et open source comme, par exemple, Framagames, un des outils de Framasoft.

    - Utiliser un faux compte en donnant des informations erronées. Nous profitons ainsi de la différence entre l'identification et l'authentification, selon la définition de Louise Merzeau (2009). De cette manière, il est toujours possible d'avoir accès aux catalogues de jeux proposés par les géants de l'industrie. Ils continuent à récolter de très nombreuses informations sur notre utilisation mais ne peuvent jamais les relier à notre identité physique. Certaines personnes peuvent cependant considérer cela comme un contournement du

[35] OpenOffice, 2019, « Why Apache OpenOffice », consulté le 26/05/2019, En ligne https://www.openoffice.org/why/

problème en nous contentant de remplacer notre identité par un pseudonyme.

- Acheter ses jeux en format physique. De cette manière, aucun compte n'est nécessaire. Il faut cependant faire attention à ne pas laisser de traces récupérables par un commerce plus traditionnel. L'inconvénient majeur est sans doute que le catalogue proposé est bien moins important. Notons également que beaucoup de jeux « physiques » ne sont en fait que de simples codes permettant de les télécharger en ligne. De plus, l'installation nécessaire à l'exploitation de ces jeux nécessite très souvent la création de divers comptes.

De cette manière, les alternatives aux grands systèmes d'exploitation sont plus nombreuses et plus faciles à exploiter que sur smartphone. Il existe en effet un catalogue assez large en fonction des préférences des utilisateurs, avec une complexité et une customisation plus ou moins importante. Concernant les logiciels, les alternatives sont elles aussi plus importantes et parfois même disponibles par défaut lors d'un changement d'OS.

Navigation et services web

Dans cette partie, nous allons trouver des solutions pour limiter nos traces numériques laissées via la navigation et l'utilisation de certains services web. Nous utiliserons cette fois encore la liste réalisée dans le chapitre précédent, à savoir le fournisseur d'accès internet, l'adresse IP, le navigateur web, le moteur de recherche et une série de services.

- Fournisseur d'accès internet : il s'agit de l'élément pour lequel il est le plus compliqué de trouver une solution. Il est même totalement possible pour ce fournisseur de savoir que nous utilisons le réseau Tor ou Tails, même si le contenu reste toujours crypté. Si l'on désire que le fournisseur ne puisse nous relier à cette activité, il faut alors jouer avec la différence entre identification et authentification. Il faudrait de cette manière utiliser un autre réseau, comme par exemple celui de quelqu'un d'autre, un cybercafé ou même un réseau pirate, comme cité précédemment pour les opérateurs téléphoniques. Notons cependant que cette attitude sort du cadre de ce mémoire, à savoir faire en sorte de ne laisser aucune trace numérique dans le cadre du monde des entreprises et non d'une surveillance d'une autre instance.

- Adresse IP : afin de cacher cette signature, il existe deux solutions très populaires, à savoir le proxy et le VPN. Un proxy est un système permettant de passer par une connexion internet indirecte et d'ainsi afficher une adresse qui n'est pas la nôtre. Cette solution ne crypte cependant pas notre trafic internet. Le VPN, pour Virtual Private Network, crée, quant à lui, un tunnel crypté entre le serveur et l'ordinateur. Notre fournisseur d'accès internet sait cependant que nous utilisons un VPN. L'utilisation du réseau Tor permet également de cacher cette adresse IP.

- Navigateur web et moteur de recherche : comme mentionné dans le chapitre précédent, il existe plusieurs extensions afin de rendre son navigateur plus sécurisé. Nous pouvons par exemple citer HTTPS Everywhere qui force un protocole https avec beaucoup de grands sites pour crypter notre communication avec celui-ci, ou Ghostery, une extension bloquant automatiquement les mouchards présents sur chaque page.

Etant donné que les extensions ne résolvent qu'une partie des problèmes, il est également possible d'opter pour d'autres navigateurs et moteurs de recherche plus respectueux des données personnelles. Pour cela, nous pouvons citer DuckDuckGo, un moteur de recherche conçu pour ne pas récupérer les traces numériques et pour protéger ses utilisateurs. Ainsi, en cas d'usage de Tails, utilisant Tor comme navigateur et DuckDuckGo comme moteur de recherche, le problème des navigateurs web et moteurs de recherche est résolu.

- Services mails : il existe de très nombreuses alternatives aux grands services mails. Nous pouvons par exemple citer Mailpile qui permet d'héberger son propre serveur ou encore Tutanota qui est également disponible sur smartphone. Le premier service voit le jour notamment grâce à une campagne de crowdfunding et le second dispose d'une version premium facultative et compte sur les dons de ses utilisateurs.

- Sites d'hébergement de vidéos : tout comme pour les services mails, il existe de nombreuses alternatives aux sites d'hébergement de vidéos. L'inconvénient majeur de ces autres plateformes est que le catalogue proposé est bien moins important, la grande majorité des vidéastes préférant utiliser les grandes plateformes. Il existe cependant une technique permettant de bénéficier des services de YouTube, le plus grand site d'hébergement de vidéos, en évitant la récupération de nos données. Il suffit en effet de remplacer le « you » de « youtube » dans l'adresse URL de n'importe quelle page ou vidéo par « hook » afin d'arriver sur le site de hooktube. Celui-ci donne accès à la totalité des vidéos YouTube mais via son propre site web et sa politique de confidentialité.

- Commerce électronique : le problème est similaire à Steam. Il existe bien de nombreuses alternatives aux grands sites mais celles-ci sont loin de garantir une meilleure politique de confidentialité et surtout loin de disposer d'un catalogue de produits aussi large que, par exemple, Amazon. Les solutions seraient alors similaires, à savoir utiliser un faux compte contenant des informations erronées ou laisser tomber cet aspect de commerce électronique pour revenir vers une version déconnectée plus traditionnelle.

- Services de streaming vidéo ou musical : tout comme pour le point précédent, les solutions principales seraient de se tourner vers l'achat du produit en format physique ou d'utiliser de faux comptes.

- Services de commande de repas : la solution la plus simple est de s'abstenir d'utiliser un tel service ou d'utiliser de fausses informations, même si la localisation doit tout de même être correcte pour garantir une bonne livraison.

- Services de cartographie en ligne : l'alternative la plus populaire est OpenStreetMap[36]. Il s'agit d'un service fournissant des données cartographiques sur le web et sur des applications mobiles. Une grande partie des services alternatifs proposant, par exemple, la création de nos propres cartes, passe via ce service. Il est enrichi par sa communauté bénévole et son utilisation est libre et gratuite. Si nous souhaitons néanmoins utiliser des sites tels que Google Maps, il est très souvent possible de désactiver certains paramètres facilitant la récupération de nos données. Par exemple, l'historique de ses positions peut être désactivé sur Google Maps. Cependant, il a été prouvé que le géant américain enregistre toujours notre position GPS malgré le paramètre désactivé. Pour empêcher cela, il faut désactiver un autre paramètre, plus discret et activé par défaut, à savoir « Activités sur le Web et les applications ». En réactivant ensuite ce dernier afin d'obtenir une explication le concernant, nous voyons alors le message « Le paramètre "Activité sur le Web et les applications" enregistre ce que vous faites sur les services, les applications et les sites Google, y compris vos recherches, vos interactions avec les partenaires Google et d'autres informations connexes telles que la position et la langue. ».

Ainsi, il existe de nombreuses alternatives aux grands services utilisés. Si cela n'est pas le cas, il est souvent possible d'utiliser un autre compte contenant de fausses informations. Afin de masquer son emprunte, beaucoup d'outils sont également disponibles, même si Tails limite à lui seul une grande partie de nos traces numériques. Le seul bémol reste cependant le fournisseur d'accès internet dont il est difficile de se passer.

[36] OpenStreetMap, 2018, « About », consulté le 16/08/2018, En ligne https://www.openstreetmap.org/about

Réseaux sociaux

Comme nous l'avons vu, les réseaux sociaux sont une source très importante de nos traces numériques via les données confiées et les données annexes. Plusieurs solutions seront présentées, toutes avec un niveau différent de réduction des traces. Nous aborderons ainsi l'éventualité d'arrêter d'utiliser un quelconque réseau et développerons plusieurs sites alternatifs.

De cette manière, la solution la plus efficace est l'arrêt total de l'utilisation de réseaux sociaux. Afin que ce changement radical ne soit pas trop handicapant, il faudrait réfléchir aux motivations qui nous poussent à utiliser ces sites. En effet, il existe certainement une ou plusieurs alternatives aux fonctionnalités qui nous intéressent. Par exemple, si nous privilégions l'aspect messagerie, il est très facile de trouver un équivalent.

Une autre solution serait d'utiliser des sites alternatifs aux grands réseaux sociaux actuels, si possible avec une politique de confidentialité respectueuse des données personnelles. La difficulté réside ici dans un élément central d'un réseau social, à savoir la présence de nombreux autres utilisateurs. Un système pourrait en effet être très performant et ne pas récupérer de données personnelles mais rester inutilisable par son nombre trop faible d'utilisateurs. Voici donc quelques alternatives intéressantes à présenter :

- Diaspora : ce site lancé en 2010 compte actuellement 646.000 utilisateurs à travers le monde[37]. Le système est basé sur trois concepts clés, à savoir la décentralisation, la liberté et la confidentialité. La décentralisation se veut en opposition aux grands réseaux sociaux qui sont propriétaires d'immenses serveurs concentrant toutes les données. Il est ici possible de choisir où nous souhaitons stocker nos données et il est même possible de créer son propre serveur. La liberté est incarnée au travers de deux éléments : le fait que Diaspora soit également un logiciel libre et qu'il n'est absolument pas demandé d'utiliser sa véritable identité. Finalement, la confidentialité signifie que nous sommes propriétaires de nos données et qu'elles ne peuvent ainsi pas être utilisées par d'autres personnes ou entités.

[37] The Federation, 2018, « Projects », consulté le 17/08/2018, En ligne https://the-federation.info/

- Mastodon : malgré son apparition beaucoup plus récente, à savoir en 2016, Mastodon compte actuellement déjà plus de 1.770.000 utilisateurs[38]. Tout comme Diaspora, le service est décentralisé et fonctionne ainsi avec un ensemble d'instances liées entre elles. Le système étant open source, chaque instance peut également être personnalisée afin de mieux correspondre aux besoins de la communauté qu'elle accueille. Pour son fonctionnement, le réseau est souvent comparé à Twitter étant donné que les utilisateurs peuvent poster des messages de maximum 500 caractères, tous classés par ordre chronologique. Il est bien entendu possible de partager des photos, des vidéos et il va également sans dire que le système ne fiche pas ses utilisateurs[39].

- Yik Yak : ce réseau social est lancé en 2013 et a la particularité d'être disponible uniquement sur smartphone lors de sa sortie. Son concept est totalement différent des deux sites précédents car il permet à chaque utilisateur de discuter anonymement avec des personnes situées dans un rayon de 8 km. Un an après sa sortie, il est déjà classé comme le 9e réseau social le plus populaire aux Etats-Unis. Le système s'offre même une version web en 2016 mais annonce sa fin en avril 2017. En effet, suite à plusieurs cas de cyberharcèlement, beaucoup d'écoles en ont interdit l'utilisation, rendant l'application bien moins populaire. Il n'existe pour le moment aucun système remplaçant Yik Yak. Certaines applications exploitent ce concept de proximité, comme Nerby et WhosHere, mais ces deux sites nécessitent la création d'un compte et diffusent de la publicité ciblée. Pour le côté anonyme, il existe le réseau social Whisper qui permet de communiquer anonymement mais sans garder ce concept de proximité. Le journal The Guardian a cependant prouvé qu'il réussit à traquer ses utilisateurs malgré l'anonymat qu'il prétend garantir[40].

[38] The Federation, 2018, « Projects », consulté le 17/08/2018, En ligne https://the-federation.info/

[39] Attilax, 2018, « Comment résister à l'espionnage permanent des GAFAM », consulté le 14/08/2018, Agora Vox, En ligne https://www.agoravox.fr/tribune-libre/article/comment-resister-a-l-espionnage-206413

[40] Lewis P. & Rushe D., 2014, « Revealed: how Whisper app tracks 'anonymous' users », consulté le 17/08/2018, The Guardian, En ligne https://www.theguardian.com/world/2014/oct/16/-sp-revealed-whisper-app-tracking-users

Nous avons donc vu que la méthode la plus efficace, mais aussi la plus radicale, serait l'arrêt total d'utilisation des réseaux sociaux. Il existe plusieurs alternatives à cette attitude extrême. Celles-ci profitent de leurs différences pour correspondre à des utilisateurs n'ayant pas les mêmes besoins et préférences.

Cadre hors web

Dans cette partie, nous allons analyser l'utilisation des informations bancaires, des cartes de fidélité et des autres services nécessitant la divulgation d'éléments d'identité.

- Informations bancaires : comme nous l'avons déjà expliqué, ces informations sont normalement confidentielles. Afin de limiter nos traces, la solution la plus efficace serait d'éviter de réaliser des paiements en ligne avec une carte bancaire. Il existe d'ailleurs des cartes prépayées achetables en format physique et utilisables sur certains sites. Une autre solution efficace, cette fois-ci pour les achats en commerce traditionnel, serait de payer en liquide. Il faut cependant bien garder à l'esprit que de telles précautions sortent du cadre de ce mémoire, à savoir en excluant une possible surveillance.
- Cartes de fidélité : la seule solution serait de ne pas utiliser ces cartes malgré les avantages financiers qu'elles proposent.
- Autres services : il s'agit cette fois d'une habitude à adopter. En effet, la seule solution légale est d'apprendre à reconnaitre ces moments où il nous est demandé de fournir des informations liées à notre identité et de les éviter. Nous voyons ainsi une limite à la réduction de nos traces quand cela nous empêche par exemple d'acheter un billet d'avion ou même un logement.

De cette manière, la méthode la plus efficace pour limiter ses traces au niveau du domaine bancaire est d'éviter le commerce électronique et de payer le plus possible en liquide. Lors d'achats en magasins physiques, l'usage d'une carte de fidélité est également à proscrire. Il faut finalement savoir reconnaitre les autres moments où il nous est demandé de fournir une information d'identité afin de s'en abstenir.

Supprimer ses traces actuelles

Comme nous l'avons vu avec les données dérivées (Schneier, 2015), certaines informations récoltées peuvent être utilisées pour en créer de nouvelles et ce sans aucune action de notre part. Nous verrons tout d'abord nos droits sur internet et nous allons nous pencher sur comment supprimer nos traces, ou du moins tenter au maximum de limiter leur impact.

Nous pouvons tout d'abord compter sur un nouvel allié, à savoir le Règlement Général sur la Protection des Données personnelles (RGPD), une loi européenne entrée en application le 25 mai 2018. Celle-ci a l'avantage de disposer d'une application extraterritoriale, à savoir contraindre les entreprises utilisant les données de résidents de l'Union Européenne même si ces organisations sont établies en dehors de l'UE. Parmi toutes les mesures de cette loi, nous pouvons notamment citer le droit d'avoir accès aux données que possède une entreprise sur nous et même le droit d'effacer certaines d'entre elles[41]. Les sites et services doivent désormais être transparents sur leur exploitation de nos données et une demande d'utilisation est même nécessaire, par exemple via une fenêtre pop-up.

Nous bénéficions ainsi de différents droits sur internet :

- Droit d'accès : il est possible de consulter toutes les informations dont dispose un organisme sur nous et de demander comment il les utilise. C'est par exemple ce droit qui est utilisé auprès de beaucoup d'entreprise pour les résultats de notre expérience.
- Droit de rectification : après consultation, il est possible de corriger ses informations personnelles et d'en supprimer certaines.
- Droit d'opposition : chaque personne a la possibilité de demander de ne pas figurer dans un fichier ou une base de données.
- Droit au déréférencement (droit à l'oubli) : il s'agit du pouvoir de demander aux moteurs de recherche de ne plus être référencé par ceux-ci ou de supprimer certains résultats affichés en introduisant notre nom dans la barre de recherche.

[41] Braun E., 2018, « Protection des données personnelles : ce qui change avec la nouvelle loi européenne », consulté le 17/08/2018, Le Figaro, En ligne http://www.lefigaro.fr/secteur/high-tech/2018/04/25/32001-20180425ARTFIG00001-le-rgpd-cette-loi-sur-les-donnees-personnelles-a-laquelle-il-faut-vous-interesser.php

Concernant ce dernier droit, le principal moteur de recherche a d'ailleurs dû mettre en ligne en 2014 un formulaire de demande de suppression des données nominatives comme résultats de recherche, afin d'automatiser ce processus. Il y a cependant certaines conditions, notamment, pour chaque lien apparaissant dans les résultats de la recherche, expliquer en quoi il est non pertinent, obsolète ou inapproprié[42].

Le déréférencement sur les moteurs de recherche n'est cependant pas suffisant pour supprimer toutes ses traces numériques. La mesure la plus efficace à mettre en place est de se débarrasser de tous les comptes qui ne sont plus utilisés, surtout s'il s'agit de réseaux sociaux. Comme nous l'avons vu, cette suppression ne se fait souvent pas instantanément et nécessite de passer par une phase de désactivation avant un effacement total et définitif.

Il faut également se pencher sur tous nos autres comptes afin de les supprimer eux aussi. La grande difficulté est qu'il est très facile d'oublier que nous nous sommes inscrits sur un site il y a très longtemps. Pour répondre à cette problématique, il existe des sites pour nous aider dans cette tâche, comme par exemple Account Killer, Just Delete Me et Knowem. Nous allons cependant nous attarder sur un autre site, à savoir Deseat.me. Cet outil retrouve tous les services sur lesquels nous nous sommes inscrits via une adresse mail Google ou Outlook. Il présente ainsi chaque site et nous demande si nous souhaitons garder ou supprimer notre compte. Une fois les éléments passés en revue, il dresse la liste des comptes dont nous souhaitons nous débarrasser. Il indique alors, pour chacun, la méthode de suppression. Certains nécessitant une demande par mail, Deseat.me peut, si nous le souhaitons, se charger d'envoyer tous les mails à notre place.[43]

Enfin, si nous n'avons pas la volonté de réaliser tout cela, il existe des agences de nettoyage numérique, soit des entreprises spécialisées dans la suppression de nos traces numériques. Cependant, celles-ci ne détiennent pas la solution miracle. Tout ce qu'elles font est donc entièrement réalisable par nous-même.

---

[42] Rees M., 2014, « Droit à l'oubli : comment effacer ses données personnelles sur Google », consulté le 17/08/2018, Next Inpact, En ligne https://www.nextinpact.com/news/87811-droit-a-l-oubli-comment-effacer-ses-donnees-personnelles-sur-google.htm

[43] Atlantico (auteur anonyme), 2016, « Effacer toute trace de votre existence sur Internet n'a jamais été aussi facile : mode d'emploi », consulté le 17/08/2018, En ligne http://www.atlantico.fr/decryptage/effacer-toute-trace-votre-existence-internet-jamais-ete-aussi-facile-mode-emploi-frederic-mouffle-2893758.html

Il existe également des entreprises spécialisées dans notre réputation, ou celle de notre entreprise, sur le web (Merzeau, 2009). Celles-ci se concentrent par exemple sur l'e-réputation sur les réseaux sociaux mais proposent également des services tels que la suppression d'avis négatifs concernant une entreprise.

Ainsi, nous disposons de différents droits sur internet, à savoir le droit d'accès, le droit de rectification, le droit d'opposition et le droit au déréférencement. Il est nécessaire de supprimer tous ses comptes qui ne sont plus utilisés et il existe de nombreux outils pour nous aider dans cette tâche. Il est également possible de se faire aider par une entreprise spécialisée.

En conclusion, nous avons vu qu'il existe la plupart du temps des solutions pour limiter ses traces numériques au travers des smartphones et ordinateurs. Il faut éviter au maximum de se créer des comptes auprès des fabricants ou des fournisseurs de services. En ce qui concerne les systèmes d'exploitation, il existe différentes alternatives libres et open source, que ce soit pour smartphone ou ordinateur. L'opérateur téléphonique et le fournisseur d'accès internet sont, quant à eux, difficiles à contourner. Pour ce qui est des applications et des logiciels, des alternatives sont très souvent disponibles et parfois même présentes par défaut en utilisant un nouvel OS. Il est également possible de profiter de nombreux services connus en étant identifié et non authentifié, c'est-à-dire en utilisant un compte contenant de fausses informations qui ne permettent pas de retrouver notre identité.

Pour les réseaux sociaux, la méthode la plus efficace serait d'arrêter totalement de les utiliser mais, afin de ne pas en arriver là, il existe des sites alternatifs. Nos traces bancaires sont limitables en payant en liquide et en ne réalisant pas de paiement en ligne. Il faut également éviter l'utilisation d'une carte de fidélité dans les commerces traditionnels.

Il reste cependant une catégorie de données, les dérivées, que nous ne limitons pas avec les méthodes précédentes. Pour celles-ci, la solution serait de supprimer ses traces actuelles ou du moins de les limiter au maximum afin de ne pas laisser la possibilité d'inférer ce type de données.

Expérience : essayer de se cacher du monde numérique

Après avoir développé toutes ces solutions pour limiter nos traces, nous allons les mettre en application en réalisant une expérience ayant comme objet central la minimisation de l'emprunte numérique d'un individu durant une certaine période. L'objectif de cette dernière est double : mieux se rendre compte des difficultés et risques d'un tel comportement, élément qui sera développé dans le chapitre suivant, et savoir s'il est possible de se cacher du monde numérique. Pour obtenir une preuve scientifique à cette expérience, nous utiliserons notre droit d'accès à nos données auprès de nombreuses entreprises pour la période durant laquelle cette expérience est réalisée. Ainsi, si des données ont tout de même été récoltées, cela validerait l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique. Par après, nous confrontons l'analyse des résultats obtenus à l'avis d'un expert avec une interview.

Cette expérience a été réalisée par moi-même entre le 10 et le 19 janvier 2019, soit pendant 10 jours. Les conditions mises en place de façon préalable sont les suivantes : suppression/désactivation des comptes sur les réseaux sociaux, abandon du smartphone, utilisation de Tails ainsi que d'une adresse mail Tutanota, absence d'utilisation des grands services en ligne pouvant récupérer les données et abstention de laisser toute autre trace, comme par exemple l'utilisation d'une carte de fidélité.

L'objectif d'une telle expérimentation est de tenter de ne laisser aucune trace tout en continuant à suivre une vie en société. C'est par exemple pour cela qu'un ancien téléphone est autorisé afin de passer des appels et envoyer des SMS et que l'accès internet est gardé. Une partie de ce mémoire est d'ailleurs rédigée dans ces conditions et, en dehors de cela, l'objectif est de poursuivre une vie ordinaire. Nous pouvons cependant considérer que de nombreux efforts personnels sont réalisés afin de limiter les traces, beaucoup de personnes n'étant par exemple pas prêtes à renoncer à leur smartphone ou disparaître des réseaux sociaux. Grâce à l'enquête réalisée, nous savons par exemple que seul un répondant sur 3 déclare pouvoir abandonner son ordinateur pendant 10 jours sans difficulté. Le postulat est que, s'il n'est pas possible de ne laisser aucune trace numérique malgré ces efforts importants, alors il n'est pas possible, dans notre société actuelle, de se cacher du monde numérique.

Le carnet de bord complet de cette expérience est disponible en annexe 3. Celui-ci contient une explication détaillée de la préparation préalable, un compte rendu quotidien et une analyse de nombreux logiciels et services utilisés ou testés durant cette période. Nous pouvons notamment citer Tails, Tor, DuckDuckGo, une suite bureautique libre, Tutanota, Hooktube, les réseaux sociaux Diaspora et Mastodon, le site de Framasoft et des logiciels libres d'édition d'images et de vidéos.

Pour obtenir les résultats de notre expérience, nous utilisons notre droit d'accès à nos données personnelles auprès de nombreuses entreprises. Nous nous servons de l'inventaire des traces numériques pour faire ressortir celles qui s'appliquent à notre situation pour ensuite réaliser nos demandes auprès des entreprises en question. L'entièreté des documents reçus est analysée et fait apparaitre les éléments suivants :

- L'opérateur téléphonique garde des informations démographiques dont certaines achetées auprès de tierces entreprises (ici, sur la solvabilité et la composition familiale). Il garde également la liste des appareils utilisés, l'activation de certains services, de plusieurs options, des préférences, etc. Parmi les éléments ayant plus d'implications dans notre expérience, l'entreprise conserve les données d'appel des 12 derniers mois, montrant que des traces ont été laissées.

- De façon très semblable, le fournisseur d'accès internet a conservé des informations démographiques très larges et « les données de trafic relatives [aux] communications ou des données de localisation », donc des traces durant la réalisation de cette expérimentation.

- Disposant de plusieurs compte Google, à savoir un personnel et un lié à un échange dans une université japonaise, une demande pour chaque compte est réalisée. L'universitaire conserve tous les messages réceptionnés (y compris les spams), donc techniquement des traces liées à notre identité étant donné que des emails ont été reçus durant l'expérience. La situation est similaire pour le compte personnel pour lequel nous pouvons également ajouter les évènements préalablement encodés dans l'agenda.

- Pour ce qui est de l'entreprise Microsoft, des données OneDrive ont été laissées, et ce malgré une absence d'utilisation du service. Il y a en effet des traces plusieurs jours à 23h59, soit certainement une sauvegarde automatique. Mis à part cela, aucune autre donnée n'a été enregistrée durant ces dix jours.

- L’entreprise Amazon conserve les communications détaillées avec le service clientèle (concernant la demande d’accès), des informations générales et de préférences, l’ensemble très détaillé des recherches réalisées, etc. Aucune trace n’a été laissée durant cette expérience.

- Concernant Facebook, bien que le compte soit désactivé, de nouvelles données ont été liées au profil durant cette période, à savoir les messages sur les discussions de groupes. Mis à part cela, rien d’autre n’a été enregistré.

- Le service Spotify garde des informations quant aux artistes et utilisateurs suivis, l’ensemble des titres dans les playlists, les informations sur l’utilisateur et l’historique d’écoute, soit, encore une fois, rien pendant notre expérimentation.

- Netflix conserve des données quant aux informations du compte, de paiement, etc. et, surtout, un historique détaillé du contenu consommé et des interactions réalisées. Il est intéressant de noter que, malgré une absence d’utilisation du service, l’entreprise a envoyé des messages (emails personnels et sur l’application) et a ainsi créé des données liées à notre compte durant l’expérience.

- Ecosia, un moteur de recherche, est le seul service à avoir répondu qu’il ne récolte et conserve aucune information quant à ses utilisateurs.

- Steam garde d’innombrables informations, comme spécifié dans l’inventaire des traces numériques, mais aucune nouvelle trace n’a été générée durant ces dix jours.

- edX conserve tous les cours suivis ainsi que des informations en rapport aux pages visitées, quand celles-ci l’ont été, etc. Aucune nouvelle trace n’a été créée.

- ASUS n’a conservé des données que lors des achats d’appareils électroniques, soit rien pendant l’expérience.

- Belfius conserve des données d’identification, de contact, de préférences de communication, les liens familiaux avec d’autres clients et des données de produits, à savoir par exemple les comptes et cartes. Aucune donnée n’a été générée durant cette expérimentation.

- Takeaway, un service de commande de repas, n’a pas pu envoyer d’informations étant donné qu’aucun compte n’est créé pour profiter du service. Le résultat est identique pour Shazam, le service de Apple, et Samsung.

Enfin, pour confirmer ces résultats, une interview de Bob De Schutter, docteur à l'université de Gand travaillant actuellement en tant que chef du Data Engineering dans une entreprise de e-commerce, est réalisée. La retranscription complète de l'interview est disponible en annexe 5. Ces données enregistrées pendant l'expérience ne le surprennent absolument pas. Selon lui, le seul moyen de ne laisser aucune trace numérique est de n'utiliser aucun appareil numérique, même le téléphone le plus basique.

En conclusion, nous avons réalisé une expérience ayant comme objet central la minimisation de l'emprunte numérique d'un individu durant une certaine période. Cependant, malgré des efforts importants de préparation préalable et durant la période de ce test, plusieurs entreprises ont généré des données, à savoir l'opérateur téléphonique, le fournisseur d'accès internet, Google, Microsoft, Facebook et Netflix. Seul un moteur de recherche a répondu qu'il ne conserve aucune donnée personnelle. D'autres services ont donné une réponse similaire mais uniquement car aucun compte n'est créé. Ces résultats ne surprennent pas Bob De Schutter pour qui le seul moyen de ne laisser aucune trace numérique est de n'utiliser aucun appareil numérique. Nous pouvons ainsi valider l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique.

Limites, difficultés et risques d'un tel comportement

Après une telle expérience, il est intéressant de se poser la question des limites de celle-ci mais également des difficultés et risques dans l'adoption du comportement d'une personne souhaitant minimiser ses traces numériques. Nous verrons pourquoi cette expérimentation n'aurait pas été tenable sur le long terme ainsi que l'impact professionnel et social de celle-ci. Nous passerons ensuite aux risques pour les entreprises, à savoir ce que cela impliquerait pour leurs revenus et leur existence.

Limites, difficultés et risques pour la personne adoptant ce comportement

Essayer de se cacher du monde numérique comporte plusieurs difficultés et risques. Nous verrons tout d'abord pourquoi cette expérience n'aurait pu être poursuivie indéfiniment et nous passerons ensuite aux risques socio-professionnels.

Ainsi, cette expérience montre ses limites pour plusieurs raisons :

- Les outils employés n'auraient pas permis de poursuivre cette expérimentation. Nous pouvons par exemple citer le peu de stockage disponible sur Tails ou encore toutes les limites des logiciels et services utilisés, éléments détaillés dans le carnet de bord en annexe 3.

- Du côté de l'UCLouvain, il aurait été nécessaire de se connecter au student corner ou au moodle à un moment ou un autre et donc laisser une trace numérique. De plus, la situation aurait été similaire avec le « Google Classroom » de l'université japonaise dans laquelle un échange a été réalisé durant les 4 mois précédant l'expérience.

- Le fait de remettre ce travail à évaluation reviendrait à laisser des traces. En effet, en plus des connexions sur diverses plateformes, des données annexes auraient sûrement été créées, par exemple avec des échanges de mails contenant des noms.

- Il aurait été impossible de réaliser le stage de fin de master. La productivité est en effet fortement diminuée par les conditions fixées pour l'expérience. La personne réalisant celle-ci est globalement plus lente et il y a d'innombrables taches qu'elle ne peut faire.

- Il est impossible de réaliser un achat en ligne, que ce soit pour des produits quotidiens mais aussi, par exemple, des billets d'avion. Or, de nos jours, il est de plus en plus difficile, voire parfois même impossible, de se séparer de ce nouveau type de commerce.

Ainsi, comme évoqué en mentionnant le stage au point précédent, un tel comportement présente des difficultés et risques, voire même une exclusion, professionnels. Il est donc impossible de réaliser un travail qualitatif pour la plupart des métiers actuels car les pratiquer implique de laisser d'innombrables traces. Nous pouvons par exemple citer le fait d'utiliser différents outils informatiques pour mener à bien son travail, se servir d'une boite mail et être

également dépendants des services mails utilisés par nos interlocuteurs, avoir son nom présent sur le site web de l'entreprise ou même sur d'autres sites, etc.

De plus, ce comportement implique de se séparer des grands réseaux sociaux alors que ceux-ci sont importants dans le monde des entreprises. Ils permettent en effet de se constituer un réseau professionnel qui est par exemple utile pour changer d'emploi, pour trouver de nouveaux clients ou simplement pour faciliter une prise de contact. Ces sites sont également utilisés par la plupart des chasseurs de têtes RH et ne pas en faire partie nous ferait manquer des opportunités professionnelles.

Ensuite, ce comportement peut mener à une auto-exclusion sociale. Fabrice Rochelandet (2010) a d'ailleurs mentionné l'importance des réseaux sociaux dans le développement d'un individu. Même s'il existe des alternatives aux grands sites, comme Diaspora et Mastodon, leur inconvénient majeur est qu'il faut que nos proches y soient également inscrits pour que ces alternatives fonctionnent. Certains pourraient argumenter qu'il n'existait aucun réseau social il y a quelques années et que tout allait pourtant bien. Cependant, ces plateformes font maintenant partie de notre environnement et s'en exclure complètement nous isolerait socialement.

De plus, un individu adoptant ce comportement se voit ajouter des difficultés dans sa propre vie mais également dans celles des personnes qui l'entourent. Par exemple, si un groupe d'amis décide quelque chose dans une discussion de groupe sur Messenger, il doit contacter le dernier membre du groupe par SMS pour avoir son avis. Il faut ainsi juste savoir que cela implique également des personnes n'ayant pas demandé à l'être.

Finalement, il est cohérent de penser qu'un tel comportement sort de la norme et peut attirer l'attention vis-à-vis d'une éventuelle surveillance. Nous pouvons d'ailleurs nous douter que cela serait catastrophique pour un état souhaitant assurer une certaine sécurité si tous ses citoyens adoptaient un tel comportement.

En conclusion, cette expérience montre ses limites car plusieurs éléments liés au statut d'étudiant et au stage de fin de master n'auraient pas permis de la poursuivre. Les outils utilisés présentent également leurs limites et certains services impliquant une récupération des données allaient devoir être utilisés. De plus, il est compliqué de poursuivre une vie professionnelle qualitative, que cela soit au niveau de la productivité ou de l'absence sur des réseaux sociaux. Il y a également des limites sociales en matière d'inclusion ou d'association de personnes nous entourant. Il est à noter qu'un tel comportement sort de la norme et peut attirer l'attention. Tous ces éléments confirment notre validation de l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique.

Risques pour les entreprises

Nous pouvons nous demander ce que cela impliquerait pour les entreprises si tous les individus adoptaient ce comportement. Pour cela, nous verrons l'impact direct sur certaines entreprises pour ensuite continuer par effet de causalité. Nous verrons l'impact sur la productivité des employés et les implications plus larges dans d'autres champs.

Dans notre expérience, nous évitons principalement d'utiliser les services de grandes entreprises. Pour savoir sur lesquelles nous pencher à titre d'exemple, nous pouvons utiliser des appellations que nous entendons souvent, à savoir le GAFAM (Google, Apple, Facebook, Amazon, Microsoft) et le FAANG (remplaçant Microsoft par Netflix) :

- Alphabet (Google) : le groupe prospère essentiellement grâce aux revenus publicitaires qui représentent par ailleurs 82% de ses entrées d'argent en 2018[44]. Avec de tels comportements de la part de leurs utilisateurs, cette source de revenus disparaîtrait totalement, ce qui mettrait l'entreprise dans une situation périlleuse. En réaction à cela, elle devrait essayer de développer des solutions alternatives, comme par exemple un système d'abonnement payant, ou se concentrer sur d'autres produits comme Android. De plus, pour que ces solutions alternatives aient une chance de subsister, l'entreprise devrait radicalement changer sa politique de confidentialité et de récolte des données. Il est également intéressant de noter que des services comme Hooktube disparaitraient. En effet, celui-ci utilise du contenu qu'il affiche sans passer de publicité et même sans comptabiliser les visionnages dans le nombre total de vues. De cette manière, sans revenus, il n'y aurait presque plus de contenu sur YouTube et le site fermerait probablement.

[44] Alphabet, 2018, « Alphabet Announces Fourth Quarter and Fiscal Year 2018 Results », consulté le 25/05/2019, En ligne https://abc.xyz/investor/static/pdf/2018Q4_alphabet_earnings_release.pdf?cache=adc3b38

- Apple : l'entreprise fonctionne de façon assez différente de Google étant donné que la majorité de ses revenus proviennent de produits physiques et non de services gratuits[45]. Face à un tel comportement, l'entreprise devrait changer la politique de confidentialité de ses produits pour que ceux-ci continuent à être achetés.

- Facebook : le réseau social ne peut pour le moment compter que sur les données de ses utilisateurs comme source de revenus grâce à la publicité[46]. Pour survivre, il devrait ainsi changer son *business model* mais surtout son fonctionnement en entier. Il serait de cette manière dans une situation critique, surtout face à des réseaux sociaux déjà établis comme Diaspora et Mastodon.

- Amazon : tout comme pour Apple, l'entreprise vend des produits physiques, ou plutôt sert d'intermédiaire entre vendeur et acheteur, et n'aurait donc pas son modèle de revenus directement impacté. Le changement majeur résiderait dans leur politique d'utilisation des données qui rendrait certainement obsolète leur système de recommandation. Cela a son importance quand nous savons que « le système de recommandation d'Amazon représenterait […] environ un tiers des revenus de l'entreprise » (Viktor Mayer-Schönberger, 2014). De plus, cela ne concernerait bien entendu pas uniquement Amazon mais bien toutes les entreprises de e-commerce. Le Dr. De Schutter nous a notifié qu'il ne faut surtout pas négliger la marge supplémentaire que font ces entreprises grâce aux données. Il est cependant plausible qu'une nouvelle forme d'e-commerce subsiste car il est parfaitement possible d'en créer une confidentielle, comme c'est par exemple le cas pour les marchés noirs en ligne.

- Microsoft : cette entreprise est fortement diversifiée dans ses produits et services. Leurs revenus dépendent à la fois de la publicité ciblée, qui n'existerait plus, et des achats directs. En changeant leur politique quant aux données personnelles et en modifiant certains de leurs *business models*, le groupe pourrait subsister.

[45] Apple, 2019, « Quarterly report pursuant to section 13 or 15(d) of the securities exchange act of 1934 », consulté le 26/05/2019, En ligne https://d18rn0p25nwr6d.cloudfront.net/CIK-0000320193/109a6139-188d-49ac-9be4-a9c78e3944e9.pdf

[46] Facebook, 2019, « Annual Report 2018 », consulté le 26/05/2019, En ligne https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/2018-Annual-Report.pdf

- Netflix : l'entreprise obtient ses revenus grâce à un système d'abonnement mensuel et ne serait ainsi pas impacté par une disparition de la publicité ciblée. Cependant, comme Amazon, l'entreprise ne pourrait plus avoir recours à son système de recommandation malgré l'importance de ce dernier. Il permet en effet d'augmenter les chances que leurs utilisateurs trouvent du contenu pertinent et qu'ils continuent ainsi à payer leur abonnement.

Nous avons de cette manière mentionné certaines entreprises mais cela concernerait bien entendu presque tous les grands groupes travaillant dans le numérique. Beaucoup d'autres entreprises dépendent de ces géants de l'informatique et seraient donc également impactées. Cela pourrait aller d'une grande multinationale au simple vidéaste qui vit de ses revenus publicitaires sur YouTube.

Plus encore, cela affecterait directement les *business models* des petites et moyennes entreprises qui, elles aussi, profitent de l'exploitation des données, comme nous l'avons vu avec Viktor Mayer-Schönberger (2014). La majorité de celles-ci seraient, d'une façon ou d'une autre, touchées par ce marché en effondrement.

Il ne faut également pas oublier que les entreprises dépendent de leurs employés. Si nous prenons ce scénario où tout le monde essaie de se cacher du monde numérique, les employés perdraient grandement en productivité, comme nous l'avons vu dans la partie précédente.

Enfin, beaucoup de sujets de recherche n'existeraient plus. Parmi ceux-là, nous pouvons bien évidemment penser à ceux orientés vers l'analyse de données. Sans intérêt économique, il y aurait bien moins de recherche malgré l'impact positif que ce domaine peut avoir dans des secteurs non-économiques, comme par exemple la médecine.

Ainsi, certains géants du numérique pourraient en partie s'adapter mais la fin de la publicité ciblée et des systèmes de recommandation se traduirait par de lourdes pertes de revenus. Une grande partie des entreprises, peu importe la taille, serait impactée, directement ou indirectement, et la productivité de leurs employés diminuerait. Cela changerait radicalement le visage de l'économie entière et nous verrions également la disparition de certains domaines tels que l'analyse de données.

En conclusion, cette expérience n'aurait pas été tenable sur le long terme et des traces auraient été laissées à un moment ou un autre. Adopter un tel comportement aurait un impact professionnel majeur en rendant impossible la réalisation d'un certain nombre de taches et la présence sur les réseaux sociaux traditionnels. Cela pourrait également mener à une auto-exclusion sociale, impliquer des personnes externes et attirer l'attention. Quant aux entreprises, elles devraient abandonner la publicité ciblée et les systèmes de recommandation. Un tel changement aurait un effet en cascade sur toutes les entreprises, d'autant plus que celles-ci seraient également impactées par une productivité en baisse de leurs employés.

Conclusion

L'objectif de ce travail est de savoir s'il est possible de se cacher du monde numérique. Dans un premier temps, cette question est précisée et un cadre est donné à ce mémoire, en expliquant par exemple certains aspects qui ne sont pas abordés. Grâce à la revue de la littérature, nous pouvons poser l'hypothèse qu'il n'est pas possible de se cacher du monde numérique, étant donné le rôle grandissant des réseaux sociaux, le *privacy paradox* et les traces générées par de tierces personnes.

Pour vérifier cette hypothèse, nous dressons un inventaire des traces numériques que nous laissons, en passant par les smartphones, les ordinateurs, certains services web, les réseaux sociaux et des éléments sortant du cadre du web. Cet inventaire est ensuite utilisé pour tenter de trouver des solutions permettant de limiter nos traces numériques. Dans cette partie, nous pouvons déjà observer que certains éléments sont particulièrement compliqués à solutionner. Par exemple, les smartphones ont une importance capitale dans notre vie alors qu'ils sont la source d'une grande part de nos traces.

Afin de mettre le point précédent en application, nous réalisons une expérience dont l'objectif est de tester, avec un individu, s'il est possible de ne laisser aucune trace numérique durant une certaine période. Avec celle-ci, le droit d'accès aux données personnelles est utilisé auprès de plusieurs entreprises afin de vérifier scientifiquement si certaines ont enregistré de nouvelles traces durant cette expérimentation. C'est le cas pour plusieurs d'entre elles, ce qui valide l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique. Une interview du docteur Bob De Schutter permet de confirmer ces résultats. De plus, une courte enquête est également élaborée, afin d'objectiver certains aspects de ce travail.

En suite logique de cette expérience, nous pouvons voir les limites, difficultés et risques du comportement qui a été adopté. Nous développons notamment pourquoi cette expérimentation n'aurait pas pu être poursuivie indéfiniment, ce qui confirme la validation de l'hypothèse selon laquelle il n'est pas possible de se cacher du monde numérique. La situation serait également fortement compliquée pour les entreprises qui seraient presque toutes impactées par de tels comportements.

Via ce travail, nous pouvons nous rendre compte de la difficulté de traiter un sujet comme les traces numériques. Plusieurs parties sont particulièrement techniques et demandent certaines connaissances en informatique. La difficulté est de passer en revue ces éléments tout en gardant à l'esprit qu'il s'agit du travail d'un ingénieur de gestion. Cependant, même si le sujet du Big Data est parfois à la limite de ce secteur d'études, ces deux domaines sont complémentaires et il est même nécessaire d'avoir une vision fédératrice les concernant.

Une autre difficulté de ce travail réside dans l'investissement consacré qui fut peut-être initialement sous-estimé, notamment à cause de l'expérience réalisée. En effet, cette dernière devait être réalisée en janvier 2019 suite à plusieurs raisons de planning. Etant donné que les parties concernant l'inventaire des traces numériques et les solutions pour les limiter devaient être écrites avant l'expérience, une part importante de ce mémoire a ainsi été écrite il y a presque un an.

Durant cette période, certains éléments tels que des services, outils et technologies ont d'ailleurs évolué, engendrant une modification de certains chapitres. Le monde numérique et ses pratiques évoluent sans cesse et ce travail sera d'ailleurs probablement obsolète dans quelques années, ce qui montre une des limites de ce mémoire.

Une autre limite concerne l'influence de la zone géographique des utilisateurs sur ce monde numérique. En effet, le rapport des individus à ce monde, ou encore sa composition même, diffèrent possiblement en fonction du lieu dans lequel nous nous trouvons. Ainsi, il pourrait être pertinent d'effectuer une comparaison entre pays ou régions du monde.

La piste d'amélioration principale pour ce travail serait par ailleurs de réaliser une expérience similaire à celle mise en place mais avec plusieurs individus. Il serait particulièrement intéressant d'obtenir les résultats et le ressenti de personnes utilisant d'autres services, n'ayant pas le même âge ou le même sexe, n'étant pas étudiant, ou encore réalisant l'expérience en ayant pu supprimer d'autres traces afin de limiter les données dérivées. Cela permettrait également d'avoir une réponse plus objective à notre question : est-t-il possible de se cacher du monde numérique ?

Enfin, il serait pertinent d'analyser plus précisément les risques pour les entreprises face au comportement adopté durant l'expérience. Il serait judicieux de réaliser une série de recherches et d'interviews de plusieurs acteurs du monde de l'entreprise et étudier leurs stratégies face à ce genre d'agissements.

Nous avons ainsi pu voir l'importance qu'a pris le monde numérique dans nos vies et son aspect incontournable. A l'avenir, d'importantes questions doivent être posées, que ce soit en matière d'évolution mais également de vie privée. En réponse à cela, différentes initiatives vont déjà dans ce sens, comme c'est par exemple le cas du RGPD. Selon Bob De Schutter, nous ne sommes actuellement qu'en train d'érafler la surface du potentiel de ce nouveau monde.

Bibliographie

Articles scientifiques et livres

Cassin B., 2009, « Google control », Cités 2009/3 (n° 39), p. 97-105, DOI 10.3917/cite.039.0097

Cecere G., Le Guel F. et Rochelandet F., 2015, « 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

De Mauro A., Greco M. et Grimaldi M., 2015, « What is big data? A consensual definition and a review of key research topics », AIP Conference Proceedings 1644 (n° 97), DOI 10.1063/1.4907823

Mayer-Schönberger V., 2014, « La révolution Big Data », Politique étrangère 2014/4 (Hiver), p. 69-81, DOI 10.3917/pe.144.0069

Merzeau L., 2009, « De la surveillance à la veille », Cités 2009/3 (n° 39), p. 67-80, DOI 10.3917/cite.039.0067

Lancelot Miltgen C., 2011, « Vie privée et marketing. Étude de la décision de fournir des données personnelles dans un cadre commercial », Réseaux 2011/3 (n° 167), p. 131-166. DOI 10.3917/res.167.0131

Lee D. et Hosanagar K., 2016, « When do Recommender Systems Work the Best? The Moderating Effects of Product Attributes and Consumer Reviews on Recommender Performance », International World Wide Web Conference 2016 (Montréal, Canada), DOI 10.1145/2872427.2882976

Nicey J., 2010, « Fabrice ROCHELANDET, Économie des données personnelles et de la vie privée », Questions de communication 18 | 2010

Rouvroy A., 2009, « Réinventer l'art d'oublier et de se faire oublier dans la société de l'information ? », Réflexions prospectives et internationales, Paris, L'Harmattan, p. 2.

Schneier B., 2015, « Data and Goliath », W. W. Norton & Company

Articles de presse, sites spécialisés et autres sources internet

Adobe Acrobat Reader DC, 2019, « Politique de confidentialité », consulté le 30/052019, En ligne https://get.adobe.com/fr/reader/

Alphabet, 2018, « Alphabet Announces Fourth Quarter and Fiscal Year 2018 Results », consulté le 25/05/2019, En ligne https://abc.xyz/investor/static/pdf/2018Q4_alphabet_earnings_release.pdf?cache=adc3b38

Apple, 2019, « Quarterly report pursuant to section 13 or 15(d) of the securities exchange act of 1934 », consulté le 26/05/2019, En ligne https://d18rn0p25nwr6d.cloudfront.net/CIK-0000320193/109a6139-188d-49ac-9be4-a9c78e3944e9.pdf

Atlantico (auteur anonyme), 2016, « Effacer toute trace de votre existence sur Internet n'a jamais été aussi facile : mode d'emploi », consulté le 17/08/2018, En ligne http://www.atlantico.fr/decryptage/effacer-toute-trace-votre-existence-internet-jamais-ete-aussi-facile-mode-emploi-frederic-mouffle-2893758.html

Attilax, 2018, « Comment résister à l'espionnage permanent des GAFAM », consulté le 14/08/2018, Agora Vox, En ligne https://www.agoravox.fr/tribune-libre/article/comment-resister-a-l-espionnage-206413

Bourguignon S., 2015, « De l'importance des réseaux sociaux professionnels », consulté le 14/01/2019, Journal Du Net, En ligne https://www.journaldunet.com/management/expert/61114/de-l-importance-des-reseaux-sociaux-professionnels.shtml

Braun E., 2018, « Protection des données personnelles : ce qui change avec la nouvelle loi européenne », consulté le 17/08/2018, Le Figaro, En ligne http://www.lefigaro.fr/secteur/high-tech/2018/04/25/32001-20180425ARTFIG00001-le-rgpd-cette-loi-sur-les-donnees-personnelles-a-laquelle-il-faut-vous-interesser.php

DTP-AG (auteur anonyme), 2017, « Télésurveillance de masse : Big Brother, une réalité en Chine », consulté le 23/08/2018, En ligne https://www.dtp-ag.com/surveillance-chine-big-brother/

Dutta A., 2019, « YouTube Business Model | How Does YouTube Make Money? », consulté le 30/05/2019, Feedough, En ligne https://www.feedough.com/youtube-business-model-how-does-youtube-make-money/

Facebook, 2019, « Annual Report 2018 », consulté le 26/05/2019, En ligne https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/2018-Annual-Report.pdf

Framasoft, 2018, « Page portail du réseau », consulté le 16/08/2018, En ligne https://framasoft.org/fr/

GNU Operating System - Free Software Foundation, 2018, « Explaining Why We Don't Endorse Other Systems », consulté le 16/08/2018, En ligne https://www.gnu.org/distros/common-distros.html

Google, 2018, « Google AdMob - Mobile App Monetization & In App Advertising », consulté le 13/08/2018, En ligne https://www.google.com/admob/

Google Play Console Help, 2018, « Transaction fees », consulté le 22/08/2018, En ligne https://support.google.com/googleplay/android-developer/answer/112622?hl=en

Guiton A., 2014, « Tails, l'outil détesté par la NSA », consulté le 16/08/2018, Le Monde, En ligne https://www.lemonde.fr/pixels/article/2014/11/20/tails-l-outil-deteste-par-la-nsa-qui-veut-democratiser-l-anonymat-en-ligne_4514650_4408996.html

Heffernan V., 2010, « Granting Anonymity », consulté le 26/05/2019, New York Times, En ligne https://www.nytimes.com/2010/12/19/magazine/19FOB-Medium-t.html?searchResultPosition=2

Hickey M., 2006, « Google Buys YouTube for $1.65 Billion », consulté le 12/08/2018, TechCrunch, En ligne https://techcrunch.com/2006/10/09/google-buys-youtube-for-165-billion/

IBM (auteur anonyme), 2018, « 10 Key Marketing Trends for 2017 », consulté le 23/07/2018, En ligne https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

Lada J., 2014, « Un système d'exploitation 100% open source est-il possible ? », consulté le 14/08/2018, PhonAndroid, En ligne http://www.phonandroid.com/systeme-exploitation-entierement-open-source-possible.html

Lapprand M., 2014 « Piratebox, ou comment créer son propre réseau Internet », consulté le 16/08/2018, Réseau International, En ligne https://reseauinternational.net/piratebox-ou-comment-creer-son-propre-reseau-internet/

Lawler R., 2018, « Netflix subscriber count hits 125 million », consulté le 12/08/2018, engadget, En ligne https://www.engadget.com/2018/04/16/netflix-subscriber-count-hits-125-million/?guccounter=1

Le Monde Informatique (auteur anonyme), 2013, « Canonical lancera Ubuntu Touch le 17 octobre », consulté le 15/08/2018, En ligne https://www.lemondeinformatique.fr/actualites/lire-canonical-lancera-ubuntu-touch-le-17-octobre-55098.html

Leloup D., 2015, « 5 systèmes d'exploitation pour snober Windows 10 (et Mac OS) », consulté le 16/08/2018, Le Monde, En ligne https://www.lemonde.fr/pixels/article/2015/08/04/cinq-systemes-d-exploitation-pour-snober-windows-10-et-mac-os_4710726_4408996.html

Lewis P. et Rushe D., 2014, « Revealed: how Whisper app tracks 'anonymous' users », consulté le 17/08/2018, The Guardian, En ligne https://www.theguardian.com/world/2014/oct/16/-sp-revealed-whisper-app-tracking-users

Martinez I., 2018, « Finding Hidden Customer Behavior Patterns Using Big Data Analytics », consulté le 25/03/2018, Dzone, En ligne https://dzone.com/articles/find-hidden-customer-behavior-patterns-using-big-d

Me and my Shadow (auteur anonyme), 2017, « Browser tracking », consulté le 12/08/2018, En ligne https://myshadow.org/browser-tracking

Me and my Shadow (auteur anonyme), 2017, « Location tracking », consulté le 03/08/2018, En ligne https://myshadow.org/location-tracking

OpenOffice, 2019, « Why Apache OpenOffice », consulté le 26/05/2019, En ligne https://www.openoffice.org/why/

OpenStreetMap, 2018, « About », consulté le 16/08/2018, En ligne https://www.openstreetmap.org/about

Panko R., 2018, « The Popularity of Google Maps: Trends in Navigation Apps in 2018 », consulté le 30/05/2019, The Manifest, En ligne https://themanifest.com/app-development/popularity-google-maps-trends-navigation-apps-2018

Pasick A., 2015, « The magic that makes Spotify's Discover Weekly playlists so damn good", consulté le 25/05/2019, Quartz, En ligne https://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/

Purism, 2019, « Librem 5 », consulté le 30/05/2019, En ligne https://shop.puri.sm/shop/librem-5/

Proximus, 2019, « Mentions légales pour les clients privés et professionnels », consulté le 26/05/2019, En ligne https://www.proximus.be/fr/id_cr_warnland/particuliers/produits/r-orphans/informations-legales.html#/tab2

Rees M., 2014, « Droit à l'oubli : comment effacer ses données personnelles sur Google », consulté le 17/08/2018, Next Inpact, En ligne https://www.nextinpact.com/news/87811-droit-a-l-oubli-comment-effacer-ses-donnees-personnelles-sur-google.htm

Silog (auteur anonyme), 2015, « Le monde numérique, c'est quoi ? », consulté le 11/01/2019, En ligne https://www.silog.fr/le-monde-numerique-cest-quoi/

StatCounter Global Stats, 2018, « Browser, OS, Search Engine including Mobile Usage Share », consulté le 12/08/2018, En ligne http://gs.statcounter.com/

StatCounter Global Stats, 2018, « Desktop Operating System Market Share Worldwide », consulté le 16/08/2018, En ligne http://gs.statcounter.com/os-market-share/desktop/worldwide

StatCounter Global Stats, 2018, « Desktop vs Mobile vs Tablet Market Share Worldwide », consulté le 12/08/2018, En ligne http://gs.statcounter.com/platform-market-share/desktop-mobile-tablet

StatCounter Global Stats, 2018, « Mobile Operating System Market Share Worldwide », consulté le 15/08/2018, En ligne http://gs.statcounter.com/os-market-share/mobile/worldwide

StatCounter Global Stats, 2018, « Social Media Stats Worldwide », consulté le 22/08/2018, En ligne http://gs.statcounter.com/social-media-stats

Steam, 2018, « Community Market FAQ / Fees », consulté le 22/08/2018, En ligne https://support.steampowered.com/kb_article.php?ref=6088-UDXM-7214#steamfee

Tamigniau M., 2016, « La fin des cartes SIM anonymes, prisées des malfrats et des terroristes, est imminente… mais elle pourrait être inutile », consulté le 26/08/2018, RTL Info, En ligne https://www.rtl.be/info/vous/temoignages/la-fin-des-cartes-sim-anonymes-prisees-des-malfrats-et-des-terroristes-est-imminente-mais-elle-pourrait-etre-inutile-821173.aspx

The Federation, 2018, « Projects », consulté le 17/08/2018, En ligne https://the-federation.info/

Annexes

Annexe 1 : Notes de lecture de Barbara Cassin, « Google Control »

Dans son ouvrage « Google Control », Barbara Cassin (2009) nous parle du géant américain de l'informatique. Elle explique son côté *Big Brother* et critique ensuite l'algorithme *PageRank* du moteur de recherche. Elle aborde la place de la publicité dans les revenus de l'entreprise, son rapport à l'opinion et sa nature, à savoir renseigner et non éduquer.

Selon elle, Google constitue une réelle menace, notamment avec son coté *Big Brother*. Le géant possède une masse considérable de renseignements grâce à ses nombreux services largement utilisés et il est même comparé à une agence de renseignement. En effet, leur mission explicite est même « d'organiser toute l'information du monde ».

Après cela, elle se penche sur l'algorithme *PageRank* du moteur de recherche. Son importance est en effet capitale car il a été prouvé que l'on se contente presque toujours des réponses situées sur la première page lors d'une recherche. Elle critique ensuite son système de *back link*, les liens qui « envoient vers », comme système de « vote » afin d'établir si une réponse est pertinente ou non. Selon elle, « être people » ne devrait pas être « l'unique critère du savoir et de la culture ».

Elle critique ensuite la place de la publicité dans les revenus de Google, ce qui peut poser des questions d'objectivité dans les algorithmes de classement. Les cofondateurs de l'entreprise se défendent cependant en assurant que les publicités sont des éléments au-dessus ou à côté du corps de page et que celles-ci peuvent servir à l'utilisateur formulant une demande précise.

L'auteure compare par après le géant américain à la sophistique. Tout comme Google, les sophistes faisaient payer leurs prestations et gagnaient de l'argent grâce à la culture et au savoir. De plus, ils ont tous deux un rapport commun à la *doxa*, à l'opinion. Or, selon Platon, « les sophistes sont nuisibles à la communauté parce qu'ils lui enseignent à se fier à l'opinion », parlant ainsi d'une coupure entre l'opinion et la vérité.

Sa dernière critique porte sur la nature même du géant américain, à savoir renseigner et non éduquer. Le moteur de recherche se contente en effet de donner les informations qu'on lui demande mais « il ne les cultive pas ».

En conclusion, Barbara Cassin (2009) réalise une critique de Google, et plus particulièrement de son moteur de recherche, sur plusieurs niveaux : son côté *Big Brother*, l'algorithme *PageRank*, l'importance de la publicité, la coupure entre opinion et vérité, sa nature et son objectif, soit renseigner et non éduquer.

Annexe 2 : Structure d'une copie des données personnelles générée par Facebook

Facebook propose un formulaire en ligne afin que chaque utilisateur puisse consulter l'ensemble des données que l'entreprise possède à son sujet. Ce formulaire permet de sélectionner une période précise, un ou plusieurs thèmes et le format, soit HTML ou JSON. Le fichier téléchargé est un dossier ZIP contenant plusieurs autres dossiers, eux-mêmes incluant de multiples fichiers. En ouvrant chaque élément, une brève explication sur celui-ci est, la plupart du temps, disponible. Dans la liste ci-dessous, un tiret fait office de dossier et un point représente un fichier. L'explication donnée est, dans la mesure du possible, celle présente dans le fichier en question.

- About you
    - Face recognition : vos paramètres de reconnaissance faciale vous permettent de choisir si vous voulez que Facebook soit capable de vous reconnaître dans des photos et vidéos.
    - Friend peer group : les étapes dans la vie de vos amis sur Facebook.
    - Your address books : coordonnées que vous avez ajoutées pour vos amis et d'autres personnes.
- Ads
    - Ads interests : vos centres d'intérêt, en fonction de votre activité sur Facebook et d'autres actions, qui nous aident à vous montrer des publicités ciblées.
    - Advertisers who uploaded a contact list with your information : annonceurs qui diffusent des publicités à l'aide d'une liste de contacts qu'ils ont importée et qui comprend les coordonnées que vous avez partagées avec eux ou avec l'un de leurs partenaires de données.
    - Advertisers you've interacted with : annonceurs dont vous avez cliqué sur les publicités sur Facebook.
- Apps abd websites
    - Apps and websites : apps auxquelles vous vous êtes connecté avec Facebook.
    - Posts from apps and websites : publications des apps auxquelles vous avez donné l'autorisation de publier en votre nom.

- Calls and messages
    - Calls logs : un journal des appels émis et reçus sur votre appareil, que vous avez choisi de partager dans les paramètres de votre appareil.
    - Message logs : un journal des messages SMS et MMS envoyés et reçus sur votre appareil, que vous avez choisi de partager dans les paramètres de votre appareil.
- Comments
    - Comments : commentaires que vous avez publiés.
- Events
    - Event invitations : invitations à des évènements que vous avez reçues.
    - Your event responses : vos réponses à des évènements, y compris Participe, Peut-être, Intéressé(e) et Ne participe pas.
    - Your events : évènements que vous avez créés.
- Files : ensemble des fichiers déposés sur des groupes.
- Following and followers
    - Unfollowed pages : pages que vous ne suivez plus.
- Friends
    - Friends : personnes qui font actuellement partie de votre réseau.
    - Received friend requests : invitations reçues d'autres personnes pour être amis sur Facebook.
    - Rejected friend requests : invitations rejetées.
    - Removed friends : personnes avec qui vous n'êtes plus connectés sur Facebook.
    - Sent friend requests : invitations envoyées à d'autres personnes pour être amis sur Facebook.

- Groups
    - Your group membership activity : dates auxquelles vous avez rejoint les groupes dont vous êtes membre.
    - Your groups : groupes dans lesquels vous êtes administrateur.
    - Your posts and comments in groups : publications et commentaires que vous avez écrits dans les groupes dont vous êtes membre.
- Likes and reactions
    - Likes on external sites : mentions J'aime sur des sites externes.
    - Pages : pages que vous avez aimées ou auxquelles vous avez réagi.
    - Posts and comments : publications et commentaires que vous avez aimés ou auxquels vous avez réagi.
- Location history : données de position GPS.
- Marketplace : activités sur la plateforme de commerce.
- Messages : ensemble de tous les messages envoyés à chaque utilisateur ou groupe.
- Network information : réseaux (affiliations avec des écoles ou des lieux de travail) auxquels vous appartenez sur Facebook.
- Other activity
    - Pokes : pokes que vous avez reçus et donnés.
    - Polls you voted on : sondages auxquels vous avez répondu.
- Pages : pages que vous administrez.
- Payment history
    - Payment history : votre historique de paiement.
- Photos and videos : ensemble des photos/vidéos publiées et des commentaires/réactions.
- Posts
    - Other people's posts to your timeline : publications que d'autres personnes ont partagées sur votre journal.

- Your posts : photos, vidéos, textes et statuts que vous avez partagés sur Facebook.

- Profile information
  - Profile information : vos nom, date de naissance, informations d'emploi et de scolarité, et lieux où vous avez vécu ajoutés sur votre profil.
  - Profile update history : historique des modifications du profil.

- Saved items
  - Your saved items : liste des publications que vous avez enregistrées.

- Search history
  - Your search history : mots, phrases ou noms que vous avez recherchés.

- Security and login information
  - Account activity : activité du compte.
  - Administrative records : données administrateur telles que des changements de mot de passe.
  - Authorized logins : les ordinateurs et téléphones mobiles que vous avez enregistrés dans votre compte Facebook.
  - Login protection data : données de protection des connexions, ensemble des cookies.
  - Logins and logout : un historique de vos connexions et déconnexions sur Facebook.
  - Used ip adresses : adresses IP utilisées.
  - Where you're logged in : périodes pendant lesquelles vous étiez activement connecté sur Facebook (contenant également les appareil).

- Your places : lieux que vous avez visités.

- Index : publications que vous avez partagées sur Facebook, publications qui sont masquées de votre journal et sondages que vous avez créés.

Annexe 3 : Carnet de bord de l'expérience

Préparation initiale

Pour préparer cette expérience, plusieurs taches ont été réalisées :

Je me suis défait de mes réseaux sociaux. J'avais ainsi déjà supprimé mon profil Twitter depuis environ 6 mois et j'en ai profité pour supprimer mon compte Snapchat que je n'utilisais presque plus. J'ai également désactivé mon profil Facebook, la suppression étant radicale vu que je comptais continuer à l'utiliser après mon expérience. Cette désactivation est cependant importante étant donné que, s'il était resté actif, des personnes auraient toujours pu mentionner mon nom quelque part, m'envoyer des messages, etc.

J'ai abandonné mon smartphone pour me tourner vers un téléphone basique, pour passer des appels et envoyer des SMS. J'ai ainsi dû vérifier si j'avais bien le numéro de téléphone de toutes les personnes que j'étais susceptible de contacter et je les ai prévenues qu'il fallait maintenant me joindre par téléphone. L'arrêt d'utilisation d'un smartphone peut sembler radical mais il s'agit de ma seule possibilité vu la difficulté d'installer un nouveau système d'exploitation comme LineageOS et le cout d'un Librem 5. De plus, j'ai remarqué que j'utilise beaucoup mon smartphone pour Facebook et Messenger, des services interdits pendant l'expérimentation.

J'ai téléchargé Tails sur une clé USB et j'ai vérifié son fonctionnement. Ma clé USB est ainsi restée branchée dans mon ordinateur durant toute la période de l'expérience. Cet outil était en effet la meilleure solution que je pouvais retenir, pour toutes les raisons déjà exprimées dans le chapitre développant les solutions pour limiter nos traces. De plus, cela m'a permis de simplement l'utiliser depuis une clé USB sans devoir installer un nouvel OS sur mon ordinateur.

Je me suis créé une adresse mail Tutanota. Cependant, comme je disposais déjà d'une adresse étudiant fournie par l'UCLouvain et d'une autre personnelle, j'ai dû me contenter d'instaurer un transfert automatique de mes mails de ces boites vers la nouvelle. L'objectif aurait été de répondre à ces messages avec cette nouvelle adresse et de demander par la même occasion à mes interlocuteurs d'utiliser celle-ci à l'avenir. Cependant, cela aurait été possible pour mon adresse personnelle mais plus compliqué pour celle universitaire. De plus, j'ai passé en revue mes derniers mails reçus pour les transférer vers ma nouvelle boite afin d'être sûr d'avoir toutes les informations dont je pourrais avoir besoin durant cette période.

Quant à l'agenda de mon téléphone, il a été remplacé par une version papier. J'ai ainsi retranscrit les événements encodés durant la période estimée de mon expérience, en laissant une certaine marge. Je suis également passé à un système de « To do » papier.

En plus de ces mesures, j'ai, pour la durée de l'expérience, abandonné tous les grands services que j'utilisais, comme par exemple les services Google ou Microsoft, le commerce en ligne, le streaming vidéo et musical, etc., soit tous les éléments que j'avais caractérisés comme pouvant laisser des traces. Je me suis également abstenu d'utiliser des cartes de fidélité et j'ai même réalisé la totalité de mes paiements en argent liquide.

L'objectif étant de tester ces conditions en essayant de poursuivre une vie classique, je me suis créé un planning avec plusieurs logiciels et sites à tester durant mon expérience. Je me suis donc contraint à utiliser les alternatives présentées dans ce travail.

Jour 1 (10 janvier 2019)

Je trouve que cette première journée s'est très bien passée. Je me suis surpris plusieurs fois à attraper mon téléphone par réflexe, afin d'y passer un peu de temps, pour tout de suite me rendre compte que je ne peux rien faire avec. Je m'y suis cependant assez rapidement habitué et cela ne m'est plus arrivé en fin de journée.

Concernant mon ordinateur, j'ai rapidement apprivoisé l'interface de Tails et son système. Je me suis cependant déjà rendu compte qu'il ne s'agit pas d'un fonctionnement vraiment tenable sur le long terme étant donné que je dois toujours travailler avec un autre stockage externe. En effet, Tails n'utilise absolument pas les disques durs et n'offre, pour ma clé USB de 8 Go sur lequel il est installé, qu'un stockage permanent de 2 Go. Je fonctionne ainsi avec une autre clé USB contenant tous mes documents concernant mon mémoire, ce qui fait que 2 ports USB de mon PC sont donc inutilisables. Je note également que Tails s'est éteint sans raison en fin de journée.

De plus, j'ai commencé à prendre pour habitude de consulter régulièrement mes mails sur mon ordinateur. En effet, j'utilisais beaucoup mon smartphone pour cela et je recevais même une notification lors de la réception d'un mail.

J’ai également utilisé les services suivants de façon significative :

- LibreOffice Writer : j’ai travaillé sur mon mémoire en utilisant ce logiciel de la suite bureautique de Tails. Malgré le fait que je doive changer mes habitudes par rapport à Microsoft Word, je trouve l’interface à la fois simple et assez complète. Je devrai cependant faire attention au format qui a légèrement changé vis-à-vis des notes en bas de pages et de la table des matières mais je trouve qu’il s’agit d’un très bon logiciel.

- Tor et DuckDuckGo : je suis également assez content de ces alternatives. Leur utilisation est intuitive et ressemble à ce dont je me servais avant l’expérience. Les seules remarques que j’ai sont que le navigateur Tor semble plus lent que ce que j’utilisais avant, caractéristique à laquelle je m’attendais, et les résultats DuckDuckGo sont légèrement moins pertinents. Cela s’explique bien entendu par le fait que ceux-ci ne sont pas dépendants d’un profil et de préférences qu’a déduit un autre moteur de recherche.

- Tutanota : j’apprécie beaucoup ce service mail et j’envisage même de le garder une fois mon expérience terminée. Il est selon moi tout aussi performant que ceux des grandes entreprises et j’apprécie son design assez épuré. Pour créer un compte, il suffit de choisir une adresse mail et un mot de passe. Aucune autre adresse et numéro de téléphone ne sont demandés. Afin de pouvoir récupérer son compte en cas d’oubli de mot de passe, un code de 64 caractères (lettres et chiffres) nous est donné lors de sa création.

- Hooktube : j’ai été assez impressionné par ce service proposant la totalité du catalogue de vidéos sur YouTube et ce sans être sur le site original. Le seul bémol du service est qu’il est plus lent que YouTube, ce qui est normal étant donné qu’il doit générer chaque page en fonction de requêtes envoyées par ses soins au service de Google. Il va également sans dire que, sans système de compte ou de *tracking*, il est moins aisé de suivre une chaine ou de trouver une vidéo particulière.

Jour 2 (11 janvier 2019)

Un élément pratique dont je ne m'étais pas vraiment rendu compte est que j'utilisais toujours l'alarme réveil de mon téléphone. J'ai ainsi dû me procurer un réveil plus traditionnel pour réaliser cette tâche. J'ai également été obligé de m'adapter pour mon jogging. J'avais en effet l'habitude de regarder préalablement mon trajet sur Google Map, pratique dont j'ai dû m'abstenir ici. De plus, j'ai toujours pris l'habitude de partir avec mon smartphone pour écouter de la musique et j'ai donc également dû m'en passer.

Concernant l'ordinateur, Tails s'est éteint sans raison en début de matinée. Mis à part cela, il n'y a pas eu de problème. Je me suis également servi du système de capture d'écran et de recoupe d'images, avec l'éditeur d'image GIMP. Une tache aussi basique fut bien entendu très simple à effectuer et j'analyserai plus en profondeur les fonctionnalités de ce logiciel dans les prochains jours. Dans les autres éléments intéressants à noter, un site m'a bloqué à cause d'un « trafic inhabituel sur mon réseau », probablement à cause de mon utilisation de Tor. Je n'ai ainsi jamais pu y accéder.

Aujourd'hui, j'ai testé le réseau social décentralisé Diaspora. Pour s'y inscrire, il faut tout d'abord choisir un « pod », soit le serveur sur lequel nos informations sont hébergées. Pour créer son compte, il est nécessaire d'avoir une adresse mail (j'en ai donné une créée spécialement pour l'occasion), un pseudonyme et un mot de passe. Une fois cela réalisé, le site nous propose d'ajouter, si nous le souhaitons, une photo de profil et des hashtags afin d'indiquer nos centres d'intérêts. Ceux-ci permettent d'afficher le contenu qui apparaît dans notre fil d'actualité. Nous avons également la possibilité de compléter un profil plus complet afin de se présenter mais rien n'est obligatoire. Etant donné qu'il s'agit d'un réseau social, je vais développer plus en détails le système de publications et de relations entre utilisateurs :

Le premier est assez complet. Il est bien entendu possible de publier du texte et le format peut même être personnalisé dans une certaine mesure (par exemple gras, italique, liste numérotée, citation, liens, etc.). Le système permet également de partager des URL, des photos, des vidéos, des sondages, sa localisation, de mentionner des personnes et d'utiliser des hashtags. Il y a deux fonctionnalités tout de même intéressantes à noter : il est possible de connecter ses comptes à d'autres réseaux sociaux afin de réaliser automatiquement des publications sur ces derniers ; un système de prévisualisation de ses publications est intégré (cf. figure 6).

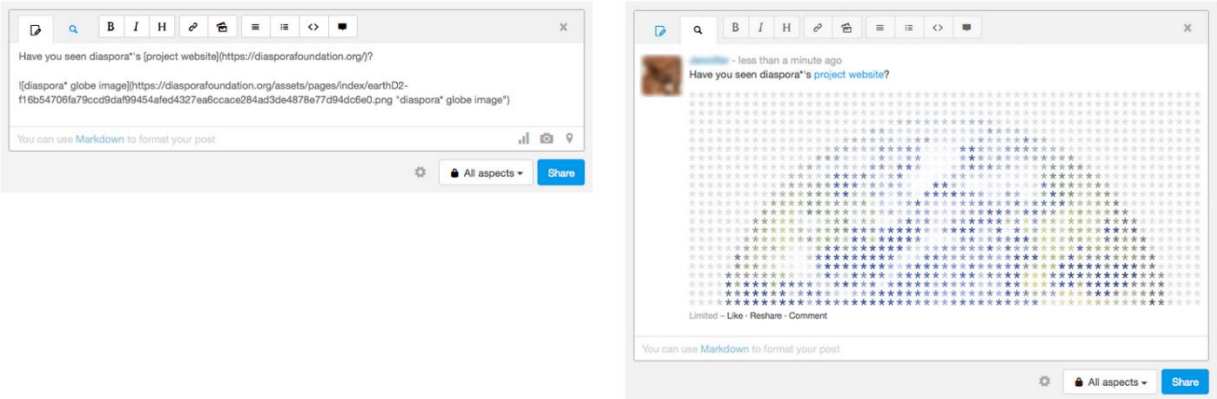

Figure 6 : Exemple de prévisualisation d'une publication sur Diaspora

La méthode pour se connecter à un autre utilisateur fonctionne différemment des grands réseaux sociaux. Il est possible de trouver des contacts en utilisant la barre de recherche, en envoyant une invitation par mail ou dans son fil d'actualité, en suivant des hashtags. Pour chaque profil, nous avons la possibilité de cliquer sur un bouton « Add contact », lui-même pouvant classer dans des catégories « Family », « Friend », « Work » ou « Acquaintances », et menant à 3 situations différentes :

- Followers : quelqu'un nous a ajouté mais cela n'est pas réciproque. Cela ne change absolument rien pour nous mais cette personne voit nos publications « publiques ».
- Following : il s'agit de la situation inverse, où nous avons ajouté une personne mais cela n'est pas le cas pour elle. Nous recevons donc ses publications publiques dans notre fil d'actualité.
- Mutual Sharing : quand nous nous sommes ajoutés mutuellement, nous voyons nos publications publiques et « limitées ». Cela dépend cependant de la catégorie dans laquelle cette personne nous a classé. Par exemple, si elle réalise une publication en cochant « Family » et « Friend » (cf. figure 7) mais nous a mis dans la catégorie « Work », nous ne voyons rien. Notons également qu'il est possible de créer nous-même nos propres catégories afin de ne pas être limité par les quatre proposées initialement.

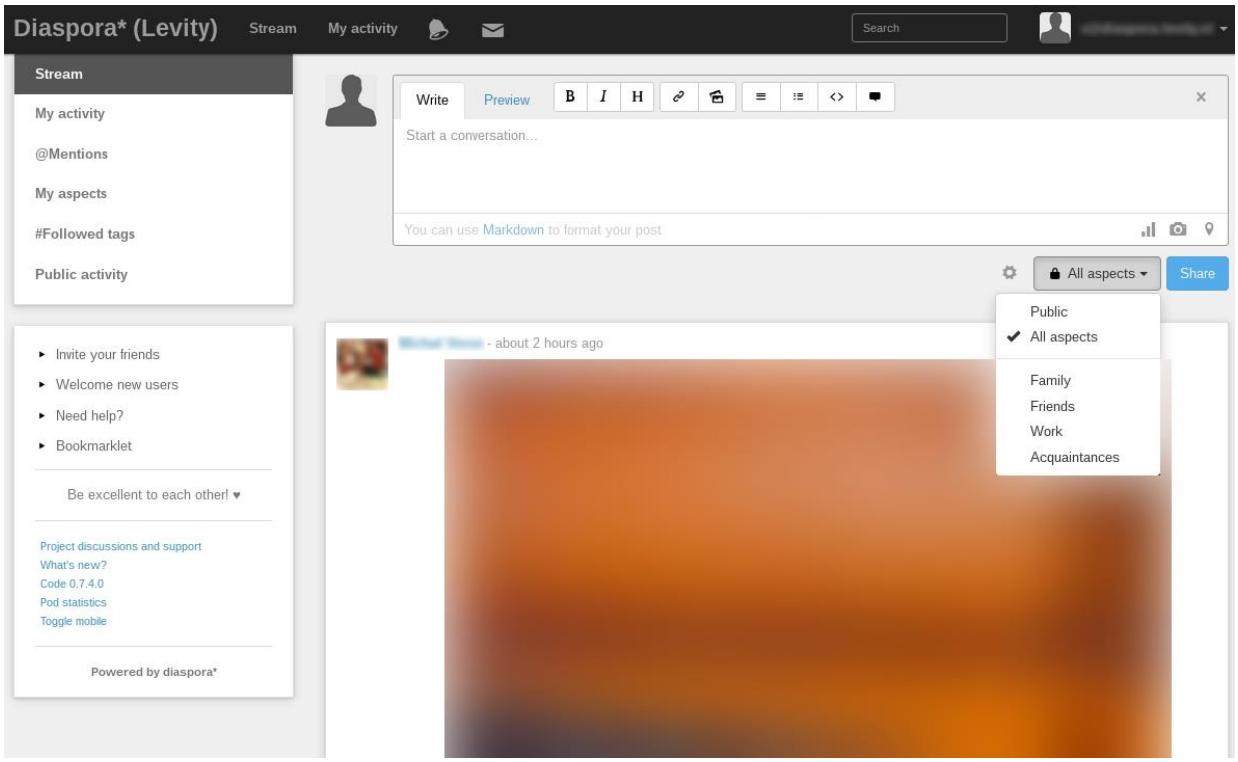

Figure 7 : Interface de Diaspora

Ainsi, ce système de relation peut sembler compliqué au début mais je m'y suis très rapidement habitué et je le trouve même supérieur à celui des réseaux sociaux plus traditionnels. Les fonctionnalités de publication sont très complètes et je précise également qu'un système de messagerie est bien entendu disponible. J'apprécie particulièrement l'interface simple à prendre en main (cf. figure 7) et je tire donc un bilan très positif du service.

Jour 3 (12 janvier 2019)

Je me suis rendu compte que je me servais très régulièrement de mon smartphone pour prendre des photos. Cela ne me posera pas trop de problèmes pour la durée de cette expérience assez courte, même si cela m'oblige à changer mes habitudes. Par contre, une alternative se serait montrée indispensable si cela avait duré plus longtemps.

Les éléments notables concernant mon ordinateur sont des problèmes de connexion internet avec Tor, m'obligeant à redémarrer Tails plusieurs fois durant cette journée. Mis à part cela, il n'y a rien de nouveau à signaler excepté un nouvel arrêt de Tails sans raison apparente.

Pour cette journée, je me suis intéressé au site Framasoft et à tous ses services offerts. Ceux-ci sont listés ci-dessous et sont accompagnés d'une brève explication.

Logiciels libres

- Framalibre : annuaire de logiciels libres.
- Framakey : service proposant une compilation de logiciels libres pour Windows, prêts à l'emploi sur une clé USB.
- FramaDVD : même service que le précédent mais sur DVD et proposant en plus de la culture libre.
- Framapack : outil permettant d'installer une collection de logiciels libres en une seule fois.

Culture libre

- Framablog : blog de Framasoft regroupant notamment divers articles.
- Framabook : collection de livres libres.
- Framabookin : bibliothèque numérique contenant des centaines d'ouvrages libres ou du domaine public.
- Framalang : communauté se chargeant de traduire collaborativement des nouvelles et livres du monde de la culture libre. Toute personne a la possibilité de s'inscrire et de participer à l'effort.

- FramaTube : site d'hébergement de vidéos.
- Framazic : site référençant une collection de musique libre, ressources que nous pouvons utiliser dans des projets.

Services libres

- Framapad : éditeur de texte collaboratif en ligne. Il est ainsi possible de créer un « pad », d'inviter des personnes et d'écrire simultanément du texte, un système de couleurs étant également mis en place.
- Framacalc : feuilles de calcul collaboratives en ligne.
- Framaform : système de formulaires et questionnaires en ligne qui se veut être une alternative à Google Forms.
- Framagenda : service de gestion et de partage de calendriers. Il est ainsi possible de gérer des agendas, des contacts et des listes de taches pouvant même être synchronisés entre ordinateurs, smartphones et tablettes.
- Framapol : service en ligne permettant de créer des sondages et pouvant par exemple servir à fixer un rendez-vous ou prendre une décision.
- Framaboard : service de gestion de taches ou de projets, en suivant la méthode Kanban.
- Framindmap : outil permettant de créer et partager des cartes mentales/heuristiques. Celles-ci peuvent notamment aider à faire un brainstorming, ordonner ses idées, apprendre et faire apprendre une leçon, réaliser des classifications ou encore identifier des éléments importants.
- Framavectoriel : logiciel permettant de créer des images vectorielles en format SVG.
- Framaslides : outil en ligne de création et d'édition de présentations.
- Framaestro : site affichant sur une même page de nombreux services web ou outils collaboratifs et permettant de travailler à plusieurs sur différents documents en même temps.

- Framabee : métamoteur de recherche regroupant les résultats d'autres moteurs de recherche mais sans conserver d'informations sur les utilisateurs.
- Framasphère : un des « pod » du réseau social Diaspora testé hier.
- Framapiaf : il s'agit d'un des serveurs de Mastodon, service similaire à Twitter dont je parlerai demain.
- Framateam, Framalistes et Framatalk : outils de discussions ayant chacun quelques spécificités.
- Framavox : service de prise de décisions collaborative permettant de discuter sur des sujets précis et de voter sur des propositions.
- Framemo : site permettant d'éditer et d'organiser collaborativement des idées sous forme de notes.
- Framanotes : outil en ligne facilitant la prise de notes et l'organisation de celles-ci.
- Framabag : service de lecture différée permettant d'extraire le contenu d'articles et de l'afficher dans une vue confortable. Il est également possible de stocker ces ressources et de les lire depuis un ordinateur, un téléphone ou une liseuse.
- Framanews : lecteur de flux RSS permettant de rester au courant de l'actualité via les flux RSS des sites de notre choix.
- Framacarte : service en ligne de création de cartes personnalisées. Il repose sur le logiciel libre uMap et sur les données de OpenStreetMap.
- Framagames : site proposant une sélection de jeux libres. Il est à noter que cette collection est pour le moment très pauvre.
- Framinetest (Edu) : jeu « bac à sable » très similaire à Minecraft visant le *gaming* de façon classique mais également l'éducation au travers d'activités pédagogiques.
- Framadrop : service en ligne permettant de partager des fichiers de façon confidentielle.
- Framabin : outil spécialisé dans le partage de textes confidentiels. Ceux-ci sont cryptés et peuvent disposer de plusieurs options comme un mot de passe, une période d'expiration au choix ou encore une destruction automatique après lecture.

- Framalink : service permettant de raccourcir des adresses URL que nous souhaitons partager.
- Framadrive : hébergement de documents en ligne. Il est à noter qu'il est pour le moment impossible de se créer un compte étant donné que le nombre maximum a été atteint.
- MyFrama : outil de partage de liens.
- FramaGit : plateforme de développement pour programmeurs semblable à GitHub.
- Framasite : service permettant la création de son site web.

Leur sélection de logiciels libres est assez importante et je suis persuadé qu'il est aisé d'y faire ses choix. Je ne m'y suis cependant pas trop intéressé car je dispose déjà de suffisamment de programmes avec Tails.

Leur collection de culture libre est pour le moment relativement limitée. Je trouve cependant l'intention très louable et j'espère que cette bibliothèque pourra s'enrichir à l'avenir.

Les services libres sont quant à eux très nombreux, certains pouvant même parfois sembler similaires sur plusieurs points. J'ai décidé d'utiliser Framagenda pour cette expérience et je me serais probablement servi de Framadrive si celui-ci avait été disponible, le stockage sur Tails étant limité. Il y a également beaucoup d'autres outils intéressants dont je ferai probablement usage même quand cette expérience sera terminée. A propos de Framagenda, le service comporte les fonctionnalités classiques d'un agenda et un système intéressant de taches et de notes. Il serait cependant encore plus utile si je pouvais profiter de sa version smartphone.

Jour 4 (13 janvier 2019)

Je me suis très peu servi de mon ordinateur aujourd'hui et je n'ai ainsi rien à signaler, mis à part un arrêt de Tails. Les seules taches que j'ai réalisées sont le test du site Mastodon et l'écriture de ce compte-rendu quotidien.

Tout comme Diaspora, Mastodon est un réseau social décentralisé nécessitant de commencer par le choix d'un serveur. Après cela, l'inscription implique d'introduire une adresse mail, un pseudonyme et un mot de passe. Le fonctionnement du site est très similaire à Twitter. Il est ainsi possible de suivre des personnes, de réaliser des publications de maximum 500 caractères, de répondre à celles-ci, de les partager et de les mettre en favori. L'interface est composée de plusieurs divisons (cf. figure 8) :

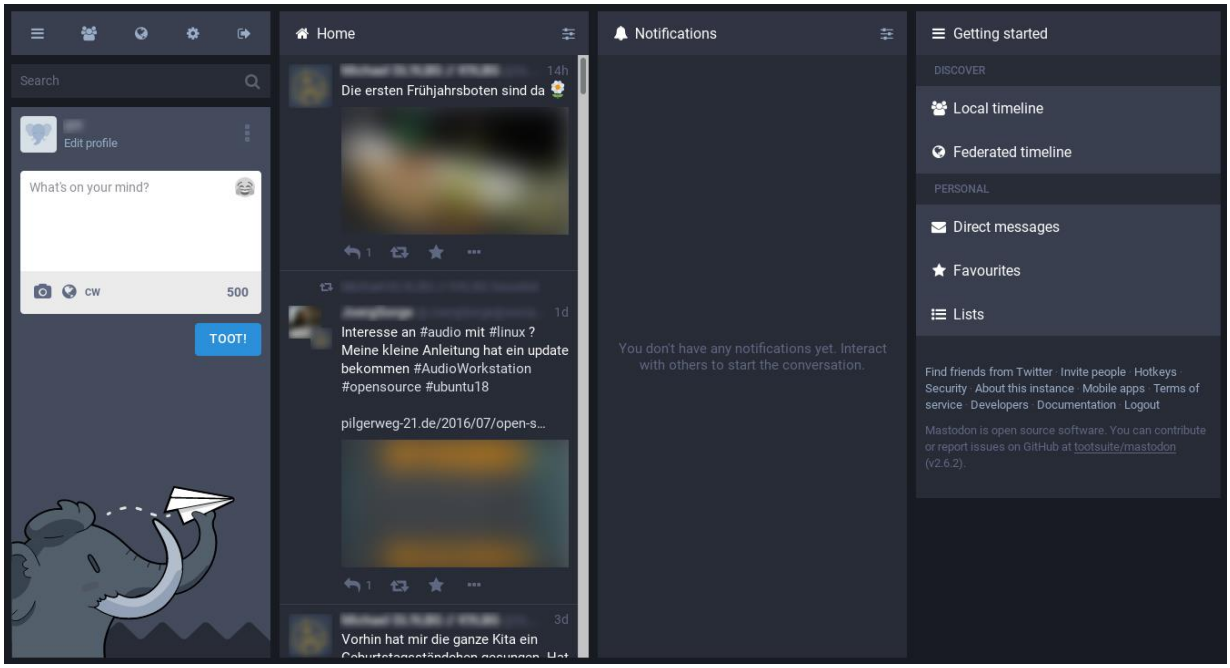

Figure 8 : Interface de Mastodon

L'élément tout à gauche permet de publier des statuts. A côté de lui, il s'agit du flux de publications personnalisé en fonction des comptes que nous suivons. A sa droite se trouvent nos notifications et, à l'extrême droite, une partie « découverte ». Celle-ci peut par exemple afficher les dernières publications du serveur sur lequel nous sommes inscrits ou celles de tous les serveurs réunis. Globalement, je trouve que Mastodon est une excellente alternative à Twitter et est très simple à prendre en main.

Jour 5 (14 janvier 2019)

Je me suis rendu compte que j'utilisais beaucoup mon smartphone en déplacement pour connaître la météo, organiser un trajet en transports en commun ou trouver un lieu avec Google Maps. Le cas de la météo fut facile à régler car je suis simplement allé la consulter au préalable depuis mon ordinateur. J'ai fonctionné de manière similaire pour planifier un trajet avec les transports en commun mais je me suis rendu compte que, s'il y avait eu un imprévu, j'aurais dû m'adapter sans compter sur mon smartphone. Enfin, le GPS de Google Maps m'était fort utile pour trouver un endroit. J'ai heureusement pu compter sur un ami qui m'accompagnait.

Pour cette journée j'ai décidé de tester la suite bureautique LibreOffice installée par défaut sur Tails. Celle-ci est composée de plusieurs logiciels :

- Writer (traitement de texte) : j'en avais déjà parlé le premier jour et je n'ai rien à ajouter, mis à part la présence de très légers *bugs* graphiques, mais rien n'entravant l'utilisation.

- Impress (présentations) : il est à la fois facile d'utilisation et complet. J'ai également essayé de modifier un power point d'une taille supérieure à 1 Go mais je n'ai jamais réussi à l'enregistrer, le logiciel refusant l'opération.

- Calc (feuilles de calcul) : après un léger temps d'adaptation, son utilisation est facile et très similaire à Microsoft Excel. J'ai n'ai cependant pas pu tester toutes les formules et il m'est donc impossible de porter un jugement quant aux fonctionnalités.

- Draw (diagrammes et graphiques) : il permet de produire des dessins, des brochures, des posters, etc. Je ne peux cependant pas le comparer étant donné que je n'ai jamais utilisé Microsoft Publisher.

- Math (éditeur de formules) : ce logiciel permet de créer des formules ou équations mathématiques et de les insérer dans nos autres documents. N'étant pas habitué à cela, je ne peux pas le comparer mais j'ai trouvé sa prise en main intuitive et agréable.

En conclusion, cette suite bureautique est une bonne alternative. Celle-ci présente cependant ses limites plus rapidement que, par exemple, celle de Microsoft, ce qui est cohérent étant donné qu'elle est gratuite. De plus, ces limites ne se perçoivent que lorsque nous poussons l'utilisation de ces logiciels et une grande partie des utilisateurs ne les atteignent ainsi pas.

Jour 6 (15 janvier 2019)

Comme chaque jour depuis le début de cette expérience, Tails s'est arrêté tout seul. Je ne l'indiquerai d'ailleurs plus dans les prochains jours étant donné que c'est un fait quotidien. Mis à part cela, il n'y a rien de nouveau à signaler.

Pour cette sixième journée, j'ai manipulé les logiciels d'édition d'images Inkscape et GIMP. Le premier permet de créer et manipuler des images vectorielles. Je ne peux ainsi pas le comparer à d'autres programmes de ce type étant donné que je n'en utilise pas mais il m'a semblé relativement simple à prendre en main. GIMP, quant à lui, est un programme plus classique que j'ai utilisé notamment pour des taches automatisées et des retouches manuelles. Je l'ai manié pendant 1h30 et j'ai trouvé son utilisation agréable. Il va sans dire qu'il présente cependant bien moins d'options que ses équivalents payants.

Jour 7 (16 janvier 2019)

Je me suis rendu compte que mon agenda sur smartphone était très pratique pour son système de rappel par notifications. En effet, ayant regardé tardivement mon agenda papier, j'ai bien failli oublier un rendez-vous.

Pour cette journée j'ai testé le logiciel de montage vidéo Pitivi. Il fut très facile à utiliser pour les manipulations très simples que je devais réaliser, à savoir assembler plusieurs fichiers vidéo et photo avec une transition entre eux. Le programme dispose d'une galerie d'effets et d'un certain nombre d'options de rendu mais il présente rapidement ses limites pour la réalisation d'un montage plus complexe. Notons que Tails s'est arrêté durant le rendu, m'obligeant à recommencer depuis le début.

Jour 8 (17 janvier 2019)

J'ai maintenant passé une semaine à essayer de me cacher du monde numérique et à tester tous les logiciels et services que je souhaitais. J'estime que j'ai maintenant un recul assez intéressant pour établir un bref bilan des conditions que j'ai fixé.

Concernant les réseaux sociaux, je n'ai pas ressenti de manque quant à l'utilisation de Facebook mais plutôt par rapport à sa fonctionnalité de messagerie Messenger. Il est en effet bien plus pratique qu'un système de SMS car il permet l'envoi aisé de fichiers informatiques et la création de discussions de groupe.

Mis à part cela, je me suis totalement habitué à ces conditions mais mon impression globale est que j'ai renoncé à beaucoup de praticité. Mon utilisation générale de l'ordinateur est en effet plus compliquée, situation similaire pour communiquer avec d'autres personnes, et mon smartphone me faisait également gagner beaucoup de temps.

Jour 9 (18 janvier 2019)

Je n'ai rien à signaler et je pense qu'il en sera de même pour la suite. Je vais poursuivre mon expérience demain et, si aucun élément nouveau n'apparait, je m'arrêterai après cette dixième journée.

Aujourd'hui, j'ai réalisé l'autre partie de ce bilan en me demandant si mon expérience serait tenable sur le long terme ou non. Les raisons qui me poussent à penser que cela n'est pas le cas sont expliquées au début du chapitre concernant les limites, difficultés et risques.

Jour 10 (19 janvier 2019)

Comme je le pensais, je n'ai rien à signaler et cette journée est ainsi la dernière de cette expérience. Je suis satisfait de celle-ci car elle m'aura, par la pratique, permis de mieux me rendre compte des limites, difficultés et risques du comportement que j'ai adopté. J'ai également eu l'occasion de tester concrètement plusieurs logiciels et services mentionnés dans ce travail. Je suis donc persuadé que cette expérimentation me permettra d'apporter une réponse éclairée à ma question de recherche.

Annexe 4 : Retranscription de l'interview du Dr. Bob De Schutter

Interview ayant eu lieu le 5 mars 2019 à 11h avec Bob De Schutter, docteur à l'université de Gand travaillant actuellement en tant que chef du Data Engineering dans une entreprise de e-commerce.

Quentin Sellier: Good morning
Bob De Schutter: Good morning, nice to meet you.
QS: Nice to meet you too. Thank you very much for accepting this interview. I am sure it will bring a lot to my master's thesis. So are you ready to answer to a few questions?
BDS: I will do my best.
QS: Is there any problem if I record this conversation? So it's easier for me later.
BDS: No no, go ahead.
QS: Ok well. So for the context, I am trying to answer the question of whether it is possible in our society to leave no digital footprint, so to become kind of invisible for companies. To answer it, I made an inventory of all the footprints that we leave, I then tried to develop a solution for each of them and I finally realized an experiment where I try myself to hide from digital world and to let no digital footprint for 10 days. So the first question is: could you say a few words about you and your experience, as you have a PhD in Physics but you work in Data Engineering?
BDS: Yes. So basically, for my studies, I studied first a bachelor's in physics. And then, I've come to do a PhD in solid state physics. For the PhD, it was focused on materials for micro components and there was a lot of Data, Data Analysis. So I wrote some programs to,…*ya*, to visualize measurements and stuff like that. I made high volumes data measurements to visualize. So in the end of my PhD, I was a bit tired of the academic world and I though "ok, what am I good at?" and it was besides doing research and physics because I didn't want to continue that anymore. So then I thought ok, it's programming, it's working with data, it's complex problems solving. And then I got the contact of the owners of vente-exclusive website and they said to me "ok when you finish your PhD, let me know" and then "we can always use a profile like you". I did it and I started as the first data scientist in vente-exclusive. So I did some schooling on myself online on data science topics, on euh…*ya*, on mobile building euh… I did extra courses at university when I was doing my PhD on machine learning and those kinds of things. I tried to find opportunities in vente-exclusive to use data science to improve

processes, to personalize things for members. So what I did actually in the beginning in vente-exclusive was my help of analyst creating Tableau dashbords. So I don't know if you know Tableau?

QS: Yeah yeah I do.

BDS: So in the beginning it was integrating some things like computing p-value, interacting with R and Python to do advanced stuff on Tableau. And then I switched to constructing the first recommender system of the website. It was a simple algorithm, a content-based recommendation algorithm, that as output it was a ranking of sales bench in our home page. Our home page is basically a collection of banners… *met* brands and basically, we tried to make an algorithm that look to the purchase history and try to order the sales with the preferences. It was a combination of getting data from big query and manage data warehouse provided by Google and then doing some calculation… then pushing the outcome of that in the cloud storage which is basically something online by Google. That was actually what I was working on and at that moment my boss at that time was the product manager for data in vente-exclusive and he went for Google and I took his place at that moment. So it was 4 years and a half not programming anymore but more taking care of the requirements, how we are using data and making sure that it works. Before he builds a new layer of using our dataledge and *voila*. And a couple month ago, we moved to a convergence with the group. We had the website in France, Italy, Spain, etc. and then I took the role of heading the data engineering for the company. So it is more leading the team of data engineers that are building data base and helping data scientists.

QS: Ok thank you. So I guess you learned actually a lot by yourself to what I understood?

BDS: I learned a lot by myself. So basically, all what I learned in term of programming. I basically had a few courses at university when I was studying physics like the beginning of object-oriented programming, some basic scripting things and, *ya*, some courses at the university but I never did a computer science degree or something tike that. But, doing my PhD, "Google is your best friend" so I started Googling around doing my own courses with whatever you could find. And, *ya*, so basically self-learning for the most part of it.

QS: Ok. So with your experience, do you think it's possible for someone to hide from digital world and to leave no digital footprint at all?

BDS: *Laughing* I think it might be possible, but I think you need to be aware of all technique that all those different things use. People usually don't know. Facebook ok, when you come on their site, but in any site where you go Facebook knows it, Google knows it. I think it might be possible but, by no way, for anyone. Also, only if you have your smartphone on. If you don't disable like everything, you will always leave digital footprints.

QS: So actually, for my experiment, I deleted or disabled all my accounts on social networks, I gave up my smartphone too, I used Tails, I don't know if you know it?

BDS: No

QS: It's like an Operating Systems that is on a USB key that immediately use Tor and everything. It was used by Edward Snowden to communicate with the journalists. I also used an email service that does not collect personal data, I did not use any of the big online services that we know like e-commerce, YouTube, etc. and even like, I didn't pay with a credit card, I didn't use any loyalty card, no GPS, all that.

BDS: Even that if you go on a website it will not track you individually, but you will still be part of the group even if you're anonymous. Just going to a website, even if they don't know who you are, you will be put in a segment or a group just based on what you do on the website. Of course they cannot track you individually but you will always be part of a digital group or a digital segment.

QS: Yeah, and also it's a lot of restrictions, what I did. When I tried to answer this question, my goal was to still be part of the society and that's why I think it's no longer possible to leave no digital footprint and still to stay in the society. Actually, after my experiment, I used GDPR to ask a lot of companies to send me all my personal data to see if they add new data from me during this period of time. I saw that, of course, I let digital footprint using my phone operator, yeah, I didn't say that but I still used an old phone that I used to do some calls and SMS but of course they got data from it. Also my internet provider, even if I used Tor, they still got information, also all the email address, and even Facebook even if it was disabled. Are you surprised of this result?

BDS: No, I'm not surprised. I think the only way which you can really have no digital footprint is to use no digital device whatsoever. Don't use a phone. As soon as you have a contract with a phone company, even if you use a Nokia 3310 for example which is like the dumbest phone, you can imagine they will still have data on you. They have cell phone towers around so even if you don't have a GPS, they still know where you make the call. I think it's really difficult. The question that I always ask is "Do you really want to have absolutely no digital footprint?" and I think your experiment is more like on the theoretical level if it's still possible.

QS: Yeah.

BDS: But I don't think it's the question that we have to ask ourselves. The question that we have to ask ourselves is "what kind of digital footprint do we want to leave and what kind of saves do we want to have on the data that companies are collecting?". I think GDPR for example is a step in the good direction. I think the reason behind is good. I think the way it was

implemented is maybe not the best way. I think it's good that people start thinking about it but you also have to be careful that you don't plant too much fear into normal people, everyday people, that are not really thinking about it because you can install a lot of fear that is sometimes not completely grounded. I think it's a difficult balance.

QS: Yeah of course

BDS: Honestly for myself, people say that "I'm afraid if I get personal advertisement". Honestly, I don't. I rather have personal advertisement than classic advertisement. You need to get advertisement anyway.

QS: Me neither actually. It was really a theorical question and, even in my introduction, I say that I won't ask the question of "do we have to do that?" or "is it good or bad?", I just ask the question of it's possible or not. Because I'm actually in Business Engineering, I have to have like a more economic side of this. So after that, I ask the consequences of this if everyone started to adopt this kind of behavior, so trying to hide. So what do you think would be the consequences for all the companies?

BDS: Oh that's a difficult one.

QS: Yeah *Laughing*

BDS: Honestly I think that for companies, even like Google,… euh… allé…people usually don't think of Google as the company that collect the most, and maybe more than Facebook. For those companies that would be a nightmare. The business model is completely focused on data. For us for example, for a company like us, I think we would feel the impact because I think you cannot underestimate the extra margin that we generate using people behavior. The website can push people to do extra purchases or to guide them to what they need so I think,…*ya*… I'm not so much an economic guy and I don't really have a business background, but I can imagine that the impact would be quite big. *Ya*, that's basically all what I have to say about it.

QS: Yeah ok. And…so the last question is: how is the situation going to evolve in the future? I mean, do you think that we will find new business models using data or did we already reached like a good level and you think it's not going to evolve?

BDS: No I think that we are only at the beginning. I think… It's euh. A lot of people are saying it but data is the gold of the 21th century and I think we're only scratching the surface. And mostly, in terms of AI, I think we cannot imagine what will be done in term of 10 years and then in 20 years. Like people are talking about self-driving cars but I think it's going to go a lot further than that. So I think it's the right time to start thinking about regulations. So it's the good time because if we don't do that, it can be out of control very easily because, allé, the

techniques in terms of AI and other things are so powerful. *pause* For dynamite, if think it's what allowed humanity to create guns and weapons but it also allowed to build things for the good so I think it's the same thing. So data and AI it's like, it's a new… how to say that in English… new… I cannot find the word…

QS: Maybe in Dutch? But I'm not sure I would know it.

BDS: *Grondstof*. It's like you have sand, or water,… it's a commodity. So it's a commodity and it can be used for good or it can be used for bad so you have to be careful. Enough regulations but not too strict so that we don't go back on innovation. But there is enough regulation so that it's hard to abuse from it and it can be used mostly for the good. One of the most important things for this decade is that we keep an eye on that.

QS: Ok… So thank you very much. It was like a quick interview but thank you very much for your time.

BDS: Ok well if you have more questions or you want more details then let me know.

QS: Ok. Thank you very much.