Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

24-May-2013

Français - Or. Anglais

DIRECTION DE LA SCIENCE, DE LA TECHNOLOGIE ET DE L'INDUSTRIE
COMITE DE LA POLITIQUE DE L'INFORMATION, DE L'INFORMATIQUE ET DES COMMUNICATIONS

Groupe de travail sur l'économie de l'information
Groupe de travail sur la sécurité de l'information et la vie privée

EXPLORATION DE L'ECONOMIE DES DONNES PERSONNELLES : REVUE DES METHODES DE MESURE DE LA VALEUR PECUNIAIRE DES DONNEES

JT03340157

DSTI/ICCP/IE/REG(2011)2/FINAL
Non classifié

Français - Or. Anglais

AVANT-PROPOS

Ce rapport est un premier examen des méthodes qui peuvent permettre de mesurer et d'estimer la valeur pécuniaire des données à caractère personnel. Les données personnelles sont de plus en plus créatrices de valeur économique et sociale, mais c'est une valeur difficile à mesurer. Cette difficulté tient non seulement au volume astronomique de données générées, mais aussi à la multitude des situations et des finalités de leur utilisation. Toute étude de la valeur des données personnelles doit commencer par une comparaison des différentes méthodes qui peuvent être utilisées pour attacher une valeur pécuniaire aux données personnelles.

En préparation des travaux dans ce domaine, le Groupe de travail sur l'économie de l'information (GTEI) et le Groupe de travail sur la sécurité de l'information et la vie privée (GTSIVP) ont organisé ensemble une table ronde sur l'économie des données personnelles et de la vie privée, qui s'est tenue le 1er décembre 2013. Trois documents de référence ont été commandés pour la table ronde, et un rapport sur les actes de cette manifestation est disponible. [1]

Le présent document a été préparé par le Secrétariat de l'OCDE (Christian Reimsbach-Kounatze, Taylor Reynolds, et Piotr Stryszowski) pour être soumis au GTEI et au GTSIVP. Il a été déclassifié par le Comité de la politique de l'information, de l'informatique et des communications (Comité PIIC) par une procédure écrite qui s'est achevée en 2013.

Il est publié sous la responsabilité du Secrétaire général de l'OCDE.

[1] www.oecd.org/sti/privacyanniversary

TABLE DES MATIÈRES

Encadrés

PRINCIPAUX RÉSULTATS

De plus en plus de modèles d'entreprise ont pour intrant central les données personnelles. Les entreprises sont maintenant capables de dériver des valeurs de marché significatives en appliquant des modèles axés sur l'utilisation productive des données personnelles, compte tenu du cadre législatif et réglementaire en vigueur. Les données personnelles se sont imposées en peu de temps comme un actif pour les processus métier, grâce à des avancées dans le domaine des TIC : pour comprendre les mécanismes économiques qui sont à l'œuvre, il apparaît nécessaire de procéder à une étude factuelle et empirique. Cette analyse permettrait aux gouvernants de mieux comprendre les enjeux de l'économie Internet, en particulier lorsqu'ils ont à élaborer les règles qui régiront la collecte et l'utilisation de ces données. Elle pose également les bases des recherches futures sur l'impact économique et sociétal de l'utilisation des données à caractère personnel.

La collecte et l'utilisation des données personnelles sont régies par les cadres réglementaires en vigueur dans les différents pays de l'OCDE. Ces cadres doivent aujourd'hui évoluer pour s'adapter au contexte : dans le monde entier, y compris dans la zone OCDE, le statut des données personnelles est en ce moment réexaminé. Comme il est noté dans le Cahier des charges de l'examen des Lignes directrices de l'OCDE sur la vie privée, le volume des données personnelles collectées, utilisées et stockées a totalement changé d'échelle, de même que la valeur des bienfaits sociétaux et économiques rendus possibles par les nouvelles technologies et les usages responsables des données personnelles. Grâce à une meilleure compréhension du rôle des données personnelles dans l'économie et dans la société, les responsables des politiques publiques seront plus au fait du contexte et mieux armés pour procéder à ce réexamen du statut des données personnelles.

Le présent rapport est une première étude des méthodes permettant de mesurer et de déterminer la valeur des données personnelles d'un point de vue purement pécuniaire (sans prendre en compte des effets indirects de l'usage de ces données sur l'économie ou la société, comme nous le verrons plus loin). Il passe donc en revue un ensemble de techniques de mesure et d'estimation et souligne les principaux avantages et inconvénients de chaque méthode. Ce rapport doit être considéré comme une première étape ; on pourra ensuite procéder à une analyse plus approfondie de la valeur de ces données du point de vue de la société et de celui des individus.

La présente analyse vise à compléter notre compréhension de tout ce qui, dans le développement de l'économie Internet, produit de l'innovation et de la croissance. L'Internet n'est pas seulement un outil efficient pour se procurer et assembler des données (notamment à caractère personnel) : c'est aussi un outil qui rend possibles une multitude d'applications émergentes novatrices qui, à partir de données, produit des services utiles aux consommateurs et aux entreprises.

Toutes les évaluations obtenues par les méthodes que nous citons doivent être utilisées avec discernement : la valeur pécuniaire dépend beaucoup du contexte, comme nous allons le voir. Par ailleurs, l'utilisation d'une seule approche risque de fausser les résultats. En outre, il n'est pas toujours facile de déterminer précisément ce qui constitue des données personnelles, ni de comparer les différentes valeurs de substitution utilisées pour les données personnelles, comme les « enregistrements » ou les « utilisateurs ». Cela étant, comprendre les valeurs obtenues par les différentes méthodes peut aider à y voir plus clair dans la gamme très étendue de valeurs pécuniaires attribuées aux données personnelles par les marchés.

L'une des méthodes permettant de déterminer la valeur pécuniaire des données personnelles pour les entreprises dont le modèle s'appuie principalement sur les données personnelles consiste à examiner la **capitalisation boursière, le chiffre d'affaires ou le bénéfice net par enregistrement**. Les évaluations

découlant de la capitalisation boursière ont une forte variabilité. Par exemple pour Facebook, entre 2006 et 2012, la valeur boursière implicite par utilisateur a fluctué entre 40 et 300 USD ; en mai 2012, elle était d'environ 112 USD. Ces fluctuations semblent être largement imputables à d'autres facteurs économiques intervenus pendant la période considérée, et pas uniquement à la valeur pécuniaire des données elles-mêmes.

Le chiffre d'affaires ou **le bénéfice net par enregistrement/par utilisateur** semblent fournir une mesure plus stable de la valeur marchande annuelle des données personnelles. Dans les cas de Facebook et d'Experian, par exemple – deux entreprises dont le modèle économique repose sur les données personnelles –, le chiffre d'affaire annuel par enregistrement/utilisateur est compris entre 4 et 7 USD. Bien qu'imprécises, ces données peuvent fournir au législateur une référence utile, même s'il est important de noter que seul le bénéfice net par enregistrement donne véritablement la mesure de la valeur économique ajoutée.

Le moyen le plus direct d'obtenir la valeur des données à caractère personnel est d'évaluer le **prix de marché** auquel sont offertes et vendues ces données lorsqu'elles le sont dans des conditions licites. Cette valeur est imprécise car elle correspond uniquement à une vente opérée dans un certain contexte à un seul acheteur, et ne reflète pas le total du « bénéfice » tiré des données sur une période prolongée. En revanche, elle fournit bien une mesure correspondant à une valeur de marché résultant de la rencontre entre l'offre et la demande. A l'heure de la rédaction du rapport, les prix pratiqués aux États-Unis pour des données personnelles étaient par exemple les suivants : 0.5 USD pour une adresse postale, 2 USD pour une date de naissance, 8 USD pour un numéro de sécurité sociale, 3 USD pour un numéro de permis de conduire, et 35 USD pour un livret militaire. Il ne s'agit là que d'estimations, mais elles ont le mérite de donner une idée de la valeur marchande relative des différentes catégories de données personnelles.

Autre approche de la valeur pécuniaire des données à caractère personnel, l'évaluation des **coûts économiques entraînés par une atteinte à la sécurité de ces données**. Il s'agit du préjudice subi par les individus dont l'identité a été usurpée et par les entreprises visées par une attaque de ce type. Les coûts de la perte et de l'utilisation malveillante des données personnelles peuvent aussi donner une idée de leur valeur, même si les chiffres recueillis sont très variés. Cette méthode ne mesure pas la valeur des données elles-mêmes mais le coût financier d'une atteinte à la sécurité des données sur une base unitaire, c'est-à-dire par enregistrement. Par exemple, lorsque la société TJX s'est fait pirater son ficher clients, elle a dû affecter 118 millions USD au titre de l'exercice 2008 pour couvrir les coûts et les pertes éventuelles (1.18 USD par enregistrement). Ces coûts n'incluent pas en revanche le préjudice subi en termes d'image, d'impact sur la marque, ainsi que les coûts d'opportunité et autres répercussions indirectes. Un exemple plus récent est celui de l'atteinte à la sécurité dont ont été victimes le réseau PlayStation Network de Sony et Sony Online Entertainment en 2011, qui a exposé au grand jour 103 millions d'enregistrements. Selon la direction de Sony, cette attaque coûtera à la société un minimum de 171 millions USD (soit 1.7 USD par enregistrement).

L'évaluation pécuniaire des données personnelles peut également passer par la réalisation de **sondages et d'expérimentations économiques** qui déterminent le prix que les entreprises auraient à payer aux individus pour qu'ils acceptent de révéler certaines des données les concernant. Ces dernières années, plusieurs expériences ont été menées pour tenter de connaître la valeur attribuée par les individus à leurs données personnelles selon les contextes. Bien que ces recherches n'en soient encore qu'à leurs débuts, elles permettent déjà de tirer deux enseignements d'ordre général. Premièrement, les individus ont tendance à faire la différence entre la valeur accordée à leurs données personnelles (c'est-à-dire la somme d'argent qu'ils jugent suffisante pour communiquer des données les concernant) et celle attachée à la confidentialité de ces données (c'est-à-dire la somme qu'ils sont disposés à payer pour empêcher que leurs données personnelles ne soient divulguées). Deuxièmement, les études empiriques montrent que

l'évaluation de la confidentialité et *l'évaluation des données personnelles* sont extrêmement sensibles au contexte.

Il existe encore une autre méthode pour déterminer la valeur économique des données personnelles d'un individu : déterminer combien cette personne serait prête à **payer pour protéger** ces données sous la forme d'une **assurance**. Cette information peut être obtenue par l'intermédiaire des polices d'assurance contre l'usurpation d'identité qui sont proposées sur le marché. Aux États-Unis, un courtier en données, Experian, propose un service de protection contre le vol d'identité – appelé ProtectMyID – pour la somme de 155 USD par an. Il n'est pas sans intérêt de comparer ce chiffre avec le chiffre d'affaires moyen par enregistrement (6.42 USD) et la capitalisation boursière par enregistrement (19.24 USD) et de regarder la différence entre les valeurs. Soulignons de plus que certains observateurs doutent de l'efficacité réelle des services censés prévenir et réparer les usurpations d'identité, et corolairement de leur validité en tant qu'instrument de mesure.

Pour comprendre la valeur économique des données personnelles, il faudrait disposer de données de meilleure qualité. Il est dans l'intérêt commun des gouvernants et des entreprises – tout au long de la chaîne de valeur – de travailler ensemble à l'efficacité des politiques relatives aux données personnelles. Les entreprises ont une bonne connaissance du nombre d'enregistrements qu'elles possèdent et ont besoin d'avoir une vision stratégique de leur valeur pécuniaire. Gouvernants et entreprises doivent coopérer plus étroitement pour mieux comprendre la valeur potentielle des données personnelles et créer des politiques publiques efficaces.

Les estimations de la valeur pécuniaire des données personnelles sont fortement dépendantes du contexte. Les effets macroéconomiques sont donc très difficiles à cerner du fait de l'absence de données homogènes et de mesures d'impacts au fil du temps. Des résultats plus intéressants pourraient être obtenus à l'aide d'études de cas mettant en évidence les impacts dans différents contextes - médecine ou transports, par exemple. Par ailleurs, la plupart des méthodes présentées sont illustrées à l'aide d'exemples de données obtenues aux États-Unis. Il convient donc d'user de circonspection pour extrapoler ces exemples aux autres pays car l'environnement économique et le cadre réglementaire peuvent avoir un impact important sur les résultats.

Même si l'on réussit à calculer la valeur pécuniaire des données personnelles, celle-ci ne traduira pas la totalité des avantages sociaux et économiques que peuvent procurer ces données – à la fois au niveau individuel et pour la société. Une analyse qui ferait abstraction du surplus du consommateur risquerait de sous-estimer les véritables avantages sociaux et économiques des données personnelles, car ces avantages ne se reflètent pas dans le prix du marché.

D'autres nouveautés qui se profilent à l'horizon pourraient aussi nous aider à appréhender la valeur pécuniaire des données à caractère personnel. Les nouveaux « coffres de données » (ou *data lockers*) permettent aux utilisateurs de consigner leur données et contrôler la manière dont elles sont partagées avec des tiers, en échange d'un pourcentage du produit de l'utilisation de leurs données. Ces bourses de données pourraient fournir de nouvelles estimations de la valeur pécuniaire des données sur le marché, et pourraient améliorer la transparence quant à la collecte, la vente et l'utilisation des données.

Introduction

Les travaux du Comité de la politique de l'information, de l'informatique et des communications (Comité PIIC) continuent de suivre les décisions actées en 2008 à Séoul lors de la Réunion ministérielle de l'OCDE sur le futur de l'économie Internet. Cette déclaration appelait l'OCDE à analyser les évolutions de l'économie Internet qui sont sources d'innovation et de croissance. Internet n'est pas uniquement un excellent outil pour se procurer et réunir des données personnelles à caractère personnel, mais indéniablement il constitue aussi une formidable plateforme pour de nombreuses applications nouvelles et novatrices qui peuvent transformer les données en services très utiles pour les consommateurs, les entreprises et les chercheurs. Les pouvoirs publics sont également à la recherche de moyens de stimuler l'innovation dans l'économie, et beaucoup des innovations qui ont connu le plus de succès depuis plusieurs années (par exemple les réseaux sociaux) sont fondées sur les données personnelles.

La Déclaration ministérielle de Séoul stipulait également que les instruments de l'OCDE, notamment des Lignes directrices de 1980 régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel (« Lignes directrices sur la vie privée ») soient réévalués à la lumière de l'évolution des technologies, des marchés et du comportement des utilisateurs, ainsi que de l'importance croissante des identités numériques, chacun de ces aspects ayant une incidence sur les données personnelles. Le mandat de révision des Lignes directrices de l'OCDE sur la vie privée appelle l'attention sur le changement d'échelle du volume des données personnelles qui sont actuellement recueillies, utilisées et stockées, ainsi que sur la valeur des avantages économiques et sociétaux rendus possibles par les nouvelles technologies, et sur les utilisations responsables de ces données.[1]

Le contexte technologique et réglementaire

À la fin des années 60 et 70, les importants progrès des technologies de l'information et de la communication ont suscité un certain émoi quant aux menaces d'atteinte à la vie privée, d'où l'apparition d'un ensemble de principes qui, aujourd'hui encore, régissent les utilisations des données personnelles. Cette période, qui a coïncidé avec l'élaboration des Lignes directrices de l'OCDE sur la vie privée, a donné lieu à l'adoption de lois sur la vie privée qui ont cours depuis lors dans les pays de l'OCDE et au-delà.

La dernière décennie a coïncidé avec d'extraordinaires avancées en ce qui concerne les capacités de stockage de données, les réseaux de transport de données à haut débit et la puissance de calcul. L'augmentation massive du volume des données personnelles et la multiplication de leurs utilisations ont été rendues possibles grâce aux progrès en termes de collecte, de stockage, d'assemblage, d'association, d'analyse et de transmission de ces données. Le passage de l'analogique au numérique a considérablement développé les possibilités de stockage et de partage des images et des vidéos. Les appareils portables permettent la collecte systématique d'informations de géolocalisation situent les individus dans le temps et dans l'espace. Les capteurs utilisés dans les secteurs de la santé, de l'environnement et de l'énergie produisent des données qui peuvent parfois être réassociées aux individus. Une grande partie de ces données sont disponibles dans le monde entier, grâce aux réseaux de communication qui permettent des flux multidirectionnels et continus de données.

Les gouvernants ont aujourd'hui engagé un travail de révision et d'actualisation des lois et des cadres relatifs à la vie privée en vigueur dans l'OCDE. Comme dans les années 60 et 70, ce sont encore les avancées technologiques – et les changements qu'ils rendent possibles eu égard au traitement des données personnelles – qui appellent aujourd'hui l'attention sur les règles et les institutions mises en place à l'époque pour encadrer la collecte et l'utilisation des données personnelles et protéger la vie privée des personnes.

Qu'entend-on par données à caractère personnel ?

L'expression qui est au cœur de l'analyse du présent rapport est celle de « données à caractère personnel ». Le rapport utilise pour cette expression la définition figurant dans les Lignes directrices de l'OCDE sur la vie privée, à savoir « toute information relative à une personne physique identifiée ou identifiable (personne concernée) ». C'est un concept vaste, qui inclut par exemple les types de données suivants :

- Contenus générés par l'utilisateur : blogs et commentaires, photos, vidéos, etc. ;
- Données liées à l'activité ou au comportement : ce que les individus recherchent et regardent sur Internet, ce qu'ils achètent en ligne, combien et comment ils paient, etc. ;
- Données sociales : contacts et amis sur les sites des réseaux sociaux ;
- Données de localisation : adresse du domicile, géolocalisation (par exemple à l'aide du téléphone portable), adresse IP, etc. ;
- Données socioéconomiques : âge, sexe, race, revenu, orientation sexuelle, affiliation politique, etc. ;
- Données officielles d'identification : nom, informations financières et numéros de comptes, informations médicales, numéro de sécurité sociale, casier judiciaire, etc.

Des recherches ont approfondi la classification des données personnelles de nombreuses façons différentes. Schneier (2010) a par exemple mis au point une taxonomie des données personnelles – en utilisant comme exemple les sites des réseaux sociaux – et établi six types de données différents : les données de service, qui sont fournies pour ouvrir un compte (par exemple : nom, adresse, informations relatives à la carte de crédit, etc.) ; les données divulguées, qui sont saisies volontairement par l'utilisateur ; les données confiées, par exemple les commentaires postés en réaction aux entrées d'autres personnes ; les données rattachées, qui concernent un utilisateur donné, mais qui ont été postées par quelqu'un d'autre ; les données de comportement, qui renseignent sur les actions effectuées par les utilisateurs lorsqu'ils se connectent à un site, et qui peuvent être utilisées pour de la publicité ciblée ; enfin, les données déduites, qui sont tirées des données publiées, du profil ou des activités d'un individu (OCDE, 2010).

Les données à caractère personnel sont également souvent classées en fonction de leur utilisation. Une distinction courante est celle qui est faite entre les données recueillies par une certaine entité en vue de leur utilisation lors de la session Internet en cours, et celles qui sont stockées pour être utilisées et analysées ultérieurement et/ou revendues à des tiers (FTC, 2009) ; on parle de cookies de session et de cookies persistants.

Il est également courant de faire la distinction entre les informations d'identification personnelle (IIP) et les informations non-IIP. Les premières sont des données qui identifient directement une personne, alors que les secondes, en principe, ne permettent pas en soi l'identification, même si la distinction entre les deux n'est pas toujours claire. Ainsi, les IIP incluent généralement le nom et l'adresse, le numéro de sécurité sociale ou autre numéro d'identification unique, les dossiers médicaux et les informations financières. Les données non-IIP désignent généralement les informations sur les termes que nous avons recherchés, les sites Web que nous avons visités, ce que nous avons acheté en ligne, le moyen de paiement utilisé, etc. D'autres types d'informations sont plus difficiles à classer (par exemple la position géographique, l'adresse IP, etc.), d'autant que les méthodes d'analyse s'améliorent et facilitent la

(ré)association d'informations qui sont elles-mêmes non-IIP mais qui, prises ensemble, peuvent permettre d'identifier les individus.

Il devient de plus en plus difficile de faire la différence entre les données personnelles et non personnelles. « Une fois qu'un lien a été établi entre une information et l'identité véritable d'une personne, toute association entre cette information et une identité virtuelle rompt l'anonymat de l'identité virtuelle. » (Narayanan et Shmatikov, 2010). Les techniques actuelles permettent souvent de retrouver l'identité d'une personne à partir des termes qu'elle recherche, des sites Web qu'elle visite, de sa géolocalisation et de son adresse IP.

Dans ce rapport, à l'instar des Lignes directrices de l'OCDE sur la vie privée, on retiendra une définition large des données personnelles, qui transcende les classifications décrites précédemment.

Objet et portée de cette étude

Comprendre la valeur économique des données à caractère personnel dans les cadres législatifs et réglementaires en vigueur est une entreprise éminemment complexe. Avec les outils et les données dont nous disposons actuellement, on ne peut réaliser qu'une analyse approximative. Le présent document s'intéresse aux méthodes utilisées pour mesurer uniquement la valeur pécuniaire de marché de ces données, et non l'ensemble de leurs impacts sociaux et économiques.

Les marchés sont les meilleurs outils dont nous disposons aujourd'hui pour connaître le prix des biens et des services. Ils ne sont toutefois pas parfaits. Si le prix du marché est déterminé par les mécanismes de l'offre et de la demande, ce prix ne tient pas toujours compte de certains avantages ou inconvénients (appelés « externalités » en économie). Pour citer un exemple, la valeur d'un espace vert correspond non seulement au prix auquel le terrain pourrait être vendu sur le marché, mais aussi aux avantages qu'en retirent les visiteurs et au renchérissement des habitations du voisinage grâce à la présence d'un parc à proximité. Par analogie, le présent rapport examine la valeur pécuniaire des données personnelles sur les marchés (ce qui correspond, dans le cas du parc, à sa valeur marchande), mais non les externalités positives et négatives résultant de leur utilisation (qui équivalent, dans l'exemple du parc, au renchérissement des biens immobiliers situés dans le voisinage du parc, au plaisir ressenti par les visiteurs et au risque de délinquance à l'intérieur du parc).

Ces externalités peuvent être importantes et ne sauraient être négligées dans l'élaboration des politiques. La valeur des données à caractère personnel doit beaucoup à leur impact socioéconomique éventuel. Un grand nombre des utilisations de ces données créent directement de la valeur sans nécessairement passer par une transaction commerciale ou être mesurables à l'aide d'une transaction, et pourtant leur impact social et économique est direct. C'est le cas par exemple de l'utilisation des données personnelles dans le domaine médical pour éviter notamment les examens redondants, les erreurs de diagnostic, etc. Si les avantages financiers sont directs et bien réels, ce n'est pas là le seul point positif. Le bienfait *personnel* de l'utilisation d'un dossier médical électronique est une amélioration du traitement du malade, dont la conséquence pécuniaire directe est indubitablement une réduction des coûts, de meilleurs résultats, etc. Il existe également des usages *socialement bénéfiques* qui créent (ou pourraient créer) de la valeur, par exemple la recherche de nouveaux médicaments, le repérage de tendances épidémiologiques ou l'amélioration des protocoles médicaux. Isoler ces externalités permettrait aux gouvernants d'avoir une meilleure appréciation de la situation lorsqu'ils établissent des règles concernant la collecte et l'utilisation des données personnelles.

Le fait de pouvoir stocker des données indéfiniment, de les associer avec d'autres et de les analyser peut être très intéressant, tant pour les personnes que pour les organisations, mais l'utilisation des données personnelles à des fins qui n'avaient pas été prévues lors de leur collecte constitue une atteinte aux

principes de base de la confidentialité. De même, les renseignements que fournit l'analyse des données concernant les inclinations, les déplacements, les centres d'intérêt et les activités des individus entraînent des risques de discrimination abusive. Ces problèmes de confidentialité – et d'autres – ne sont pas abordés dans le présent rapport ; idem pour les coûts économiques que peuvent entraîner des utilisations irresponsables ou illégales des données personnelles pour les individus, les organisations et les sociétés tout entières.

L'étude des seules méthodes d'évaluation pécuniaire ne nous renseigne donc pas sur la totalité des avantages et des coûts sociaux et économiques des données personnelles, mais elle permet en revanche de jeter des bases importantes en prévision de travaux ultérieurs. Toute analyse visant à appréhender les impacts sociaux et économiques dans leur globalité devra commencer par une étude des marchés sur lesquels les biens sont échangés.

Pour la suite, ou pourrait envisager de définir un cadre conceptuel pour décrire comment les données personnelles peuvent créer de la valeur économique et sociale, et de construire des indicateurs pour déterminer cette valeur. Il faudrait pour cela étudier des secteurs autres que la publicité et les médias sociaux (comme le secteur médical, la finance, l'industrie automobile, etc.), et voir comment chacun de ces secteurs exploite déjà les renseignements obtenus grâce à l'analyse des données - renseignements issus de données internes collectées par l'organisation elle-même pour travailler plus efficacement, de données obtenues par l'intermédiaire des échanges commerciaux de données, ou dérivés en combinant des données provenant de différentes sources, etc. Cela permettrait peut-être en outre de mettre en évidence d'autres indicateurs de la croissance économique qui pourrait résulter de la valeur des données personnelles.

Les valeurs des données personnelles obtenues grâce aux différentes méthodologies sont, faut-il le rappeler, imprécises, mais elles représentent une première étape dans la compréhension du large éventail de valeurs commerciales que peut avoir cet actif immatériel qui prend de plus en plus d'importance.

L'économie des données personnelles : cartographie des chaînes de valeur et des modèles économiques

L'examen des chaînes de valeur permet de comprendre le rôle des différents acteurs et de repérer les domaines d'activité dans lesquels les données sont mieux valorisées. L'approche de la question par le biais de la chaîne de valeur peut aussi offrir une vue d'ensemble de l'activité et permettre de repérer des domaines auxquels les gouvernants voudront peut-être s'intéresser davantage.

Dans cette section, nous nous pencherons sur les chaînes de valeur des données personnelles aux différentes étapes du cycle de vie des données, depuis leur collecte jusqu'à leur utilisation, et sur les modèles d'entreprise déployés pour générer et extraire la valeur de ces données au long de leur cycle de vie. On s'intéressera en particulier aux innovations rendues possibles par les données personnelles : produits et services nouveaux, gains d'efficience dans les processus, nouvelles formes d'analyse et création de valeur pour les entreprises comme pour les utilisateurs. Ces chaînes de valeur présentées en contexte, pour mettre en lumière la manière dont les données personnelles sont utilisées selon les secteurs. Les cadres en matière de vie privée peuvent avoir un impact sur la chaîne de valeur et peuvent varier selon les pays, voire selon les secteurs au sein d'un même pays.

Il existe différents modes de collecte des données personnelles :

- Les données peuvent être *fournies* ou *consenties* par les individus qui partagent de manière explicite des informations sur eux-mêmes ou sur autrui (par exemple lorsqu'ils créent un profil sur un réseau social, saisissent les coordonnées de leur carte bancaire pour acheter en ligne,

fournissent leurs informations personnelles pour s'inscrire à un service en ligne, ou postent des renseignements concernant des amis, collègues, membres de leur famille, etc.) ;

- Les données peuvent être *observées de manière licite*, c'est-à-dire captées en enregistrant les activités des utilisateurs, ce qui diffère des données fournies volontairement (par exemple, les préférences de navigation sur Internet, les données de géolocalisation lors de l'utilisation d'un téléphone portable, ou le comportement téléphonique de l'utilisateur) ;

- Les données peuvent être *déduites* de l'analyse des informations personnelles (ainsi, la solvabilité d'une personne peut être calculée à l'aide d'un certain nombre de facteurs se rapportant à ses antécédents financiers). Elles peuvent également être *déduites* par recoupement de plusieurs informations en apparence anonymes (Narayanan et Shmatikov, 2010).

Dans chacun de ces modes de collecte (données fournies, consenties, observées ou déduites), les données personnelles sont tout d'abord recueillies ou consultées, puis stockées, assemblées, traitées et enfin utilisées et analysées. Chacune de ces étapes présente des caractéristiques particulières et peut faire intervenir des acteurs différents (graphique 1). Le cycle de vie des données personnelles peut s'envisager comme une chaîne de valeur composée des quatre étapes[2] suivantes :

- Collecte/Consultation

- Stockage et assemblage

- Analyse et distribution

- Utilisation

Graphique 1. Chaîne de valeur des données personnelles

Données personnelles	Collecte/Consultation	Stockage et assemblage	Analyse et distribution	Utilisation
Fournies intentionnellement (par exemple indiquer ses hobbies, centres d'intérêt, préférences, domaines de compétence, etc.)	• Téléphone mobile • Blogs et forums de discussion • Réseaux sociaux, professionnels et thématiques • Contenus générés par l'utilisateur • Programmes de fidélité gérés par les commerçants • Appareils intelligents • Applications • Capteurs • Etc.	• FAI et opérateurs de téléphonie • Organismes publics (centres des impôts, cadastre, etc.) • Réseaux sociaux en ligne • Établissements financiers • Organisations de santé • Services de réseau • Commerces • Etc.	• Commerces et services • Administration publique • Établissements financiers • Prestataires de soins médicaux • Entreprises de publicité et marketing en ligne • Analystes et courtiers en données • Etc.	• Entreprises • Administrations et organismes publics • Utilisateurs finals
Observées (par exemple, géolocalisation, historique de navigation, habitudes d'achat, etc.)				
Déduites (ex : solvabilité, profil établi à partir des activités en ligne, etc.)				

Source : OCDE, à partir de WEF (2011).

Différents acteurs interviennent tout au long de la chaîne de valeur (particuliers, entreprises, organismes publics, organisations à but non lucratif, etc.). Certains acteurs ne sont présents à toutes les étapes. Ainsi, les courtiers en données n'*utilisent* pas habituellement les données personnelles eux-mêmes, mais ils en font un *traitement* et les *revendent*. Il arrive au contraire qu'un même acteur prenne en charge les données à toutes les étapes de la chaîne de valeur. C'est le cas par exemple d'une compagnie aérienne ou d'un commerçant, qui recueille des données personnelles dans le cadre d'un programme de fidélisation des clients, les stocke puis les assemble, et enfin les traite et les utilise dans son propre modèle d'entreprise (par exemple en proposant à certains clients des offres ciblées).

Enfin, il est intéressant de noter que la délimitation entre les données fournies, consenties, observées ou déduites n'est pas très claire. Les données fournies, consenties et observées déterminent fortement les données déduites, et celles qui sont fournies et consenties sont souvent inexactes.

Collecte / accès

Première étape de la chaîne de valeur : la collecte et l'accès aux données personnelles, conformément au cadre juridique applicable. Ce processus se retrouve dans tous les secteurs de l'économie et les données proviennent d'une multitude de sources. Avec la croissance d'Internet et le développement de technologies

de communication génératrices d'énormes quantités de données, cette étape a connu une profonde mutation. Les opérateurs de réseaux mobiles disposent par exemple d'informations détaillées sur les personnes, y compris leur localisation (via les réseaux de téléphonie mobile) et leur journal d'appels. Les fournisseurs d'accès Internet (FAI) ont quant à eux potentiellement accès à des informations très précises sur les habitudes d'utilisation de leurs abonnés. Même dans les secteurs autres que les TIC, les commerçants disposent désormais d'outils beaucoup plus performants pour suivre les achats réalisés par les clients à l'aide de leurs cartes de fidélité, voire en utilisant des capteurs installés dans les magasins. Cette section fournira plusieurs exemples de la collecte de données personnelles.

La collecte directe de données s'effectue notamment via l'enregistrement des achats et les programmes de fidélisation des clients gérés par les commerces, les services de transport et l'hôtellerie-restauration. Autres sources de données : les dispositifs d'identification de l'utilisateur, de connexion et d'authentification de la transaction mis en œuvre par toute une série d'opérateurs – fournisseurs de téléphonie et FAI, notamment – les contenus générés par les utilisateurs - courrier électronique, commentaires postés sur des blogs et des forums de discussion, sites Web personnels et professionnels et de téléchargement de photos et vidéos – et enfin la participation à des réseaux en ligne (sociaux, professionnels ou thématiques).

Les programmes de fidélité des commerçants et des fournisseurs de services permettent d'établir un profil des clients et ainsi de mieux les comprendre, de réduire le coût des transactions pour les deux parties, ainsi que de proposer des réductions et des offres spéciales pour les clients fidèles. Les fichiers-journaux qui enregistrent les connexions et les autres transactions fournissent des informations sur le comportement et les besoins des utilisateurs, ce qui permet aux prestataires de fournir des services plus adaptés. La pléthore de contenus générés par l'utilisateur depuis quelques années multiplie également les possibilités de profilage et de ciblage de certains individus, à qui sont proposés des produits et services susceptibles de les intéresser.

Les données peuvent aussi être recueillies de manière plus indirecte, notamment dans de nouveaux modèles économiques, où la collecte de données va prendre différentes formes : localisation d'un utilisateur de téléphone portable ou informations de localisation Internet, dans le but de fournir des contenus et des services susceptibles d'intéresser l'utilisateur en fonction de l'endroit où il se trouve, ou de limiter l'accès à des contenus se rapportant à un certain périmètre géographique. Les autres sources d'observation des données personnelles sont notamment l'historique de navigation, les pages visitées ou les téléchargements effectués, qui permettent de suivre les activités en ligne des utilisateurs. Autres données qui peuvent être collectées, les achats et les transactions, les accès à certaines applications spécifiques et leur utilisation – applications bureautiques et domotiques, réseaux électriques intelligents et sécurité (par exemple la télésurveillance, les capteurs, etc.). Les technologies réseau permettent désormais de suivre les individus et les objets dans un nombre croissant d'activités, d'où l'apparition de nouveaux modèles économiques. Beaucoup de ces modèles servent essentiellement à mieux connaître les consommateurs afin de leur proposer des produits et des services adaptés, à réduire les coûts pour les utilisateurs en termes de recherche des utilisateurs et de coûts de transactions, ainsi qu'à accroître l'efficience des fournisseurs et des prestataires, qu'ils soient privés ou publics.

Enfin, les données peuvent être non pas recueillies, mais générées par l'analyse et peuvent aboutir à l'établissement de profils de préférences des individus d'après leur activité en ligne, qui pourront être utilisés pour des publicités et des offres spéciales ciblées ; des notes solvabilité peuvent être générées ; les individus peuvent être repérés et suivis grâce à leur géolocalisation téléphonique et/ou Internet, à des capteurs, à des systèmes de domotique ou bureautique, à des réseaux électriques intelligents, etc. Nous allons maintenant examiner ces activités de traitement et d'analyse des données à caractère personnel.

Stockage et agrégation

Une fois recueillies, les données peuvent être stockées et agrégées. C'est la deuxième étape de la chaîne de valeur. Les différents éléments de données sont organisés et classés en vue de leur traitement et leur analyse ultérieurs.

De nombreux enregistrements de données personnelles (tels que les coordonnées, les informations de la carte de paiement ou de crédit, les données de compte et d'authentification de connexion) sont sauvegardés par différents prestataires de services : FAI et opérateurs de téléphonie, commerçants, sociétés de transport, professionnels de la santé, services de réseaux et organismes gouvernementaux. Les contenus générés et postés par les utilisateurs sont également conservés par tout un ensemble de fournisseurs de services et de contenus : réseaux sociaux et professionnels, réseaux sociaux thématiques, blogs, sites de partage de photos et de vidéos, et services de messagerie. Par souci d'économie, les données se rapportant aux individus et aux organisations sont de plus en plus stockées à distance et consultées en ligne.

Beaucoup d'autres intervenants de la chaîne de valeur stockent et agrègent des données. Les FAI et de services de recherche conservent les historiques de navigation et les enregistrements de suivi et de recherches Internet ; les professionnels de la santé et toute une série d'organismes publics et privés – assurances santé et employeurs notamment – conservent des données médicales ; les commerçants, banques et autres établissements financiers, les employeurs et les administrations fiscales conservent des données bancaires ; enfin, les opérateurs de téléphonie mobile, les FAI, les services de réseau et les sociétés de transport, entre autres, stockent des informations de géolocalisation.

Analyse et distribution

La troisième étape de la chaîne de valeur des données personnelles consiste à prendre les données recueillies et stockées et à les combiner avec d'autres informations afin d'établir des profils et des enregistrements détaillés ainsi que de dégager des tendances au niveau macro, pour ensuite les utiliser à des fins diverses. L'essentielle de la valeur qui est ajoutée à ce stade provient de la fusion de données provenant de différentes sources pour créer un profil et de l'analyse de ces données, ce qui permet de déduire des informations que l'on ne pourrait peut-être pas obtenir autrement. Les sources peuvent être des ensembles de données librement accessibles, des données propriétaires détenues par les entreprises et des informations émanant d'établissements de recherche. Cette agrégation de données offre des possibilités inédites d'observation, ce qui crée de formidables débouchés pour de nouveaux produits et services. Précisons que les données peuvent subir plusieurs cycles d'analyse et de distribution, et que de nouvelles données peuvent être ajoutées à chaque itération. Lorsque les enseignements tirés sont utilisés pour affiner le profil des utilisateurs, il est courant que les sociétés d'analyse de données revendent sur le marché les profils issus de cette opération : ces acteurs constituent donc une source d'information intéressante sur la valeur pécuniaire des données personnelles. D'autres types de enseignements peuvent être tirés de ces analyses : elles peuvent servir à améliorer le service client et la qualité des produits, à suivre les problèmes d'interactions médicamenteuses ainsi que les courbes quotidiennes de trafic.

Ce travail est souvent réalisé par des entreprises dotées d'une infrastructure avancée, de capacités d'analyse et de réseaux de distribution bien développés. Il peut s'agir des intervenants traditionnels de la chaîne de valeur ou de nouveaux acteurs apparus pour remplir de nouveaux besoins ou profiter de nouvelles possibilités. Les acteurs traditionnels du traitement des données personnelles sont notamment les commerçants et les prestataires de services utilisant un logiciel de gestion des relations clients (CRM), des systèmes de veille stratégique et des programmes de fidélité. Les acteurs récents sont par exemple les analystes et courtiers en données spécialisées, les sociétés d'études de marché et de publicité en ligne. Autour du « pistage » (ou tracking), s'est développée une nouvelle branche d'activité qui suscite une vague d'innovations dans le domaine de la publicité (voir l'encadré 1).

Encadré 1. Analyse, distribution et utilisation des données à caractère personnel : l'exemple de la publicité en ligne

La publicité en ligne fait intervenir différents acteurs : les éditeurs de sites web, les annonceurs, et les réseaux publicitaires intermédiaires, qui mettent en relation éditeurs et annonceurs souhaitant toucher un public sur Internet. Les réseaux publicitaires ou « ad-networks » jouent un rôle central. Il consiste à vendre des espaces pour le compte des éditeurs Web ; on les qualifie souvent de « régies publicitaires externes » car ils travaillent avec différents éditeurs partenaires, auxquels ils achètent de l'espace publicitaire disponible pour le revendre à des annonceurs. La relation est utile pour les deux parties : pour les éditeurs Web c'est une manière de rentabiliser leurs contenus sans avoir à percevoir d'abonnements ou autres droits d'utilisation, et pour les annonceurs c'est un moyen de toucher le public qu'ils veulent atteindre. Ces intermédiaires sont particulièrement importants pour les petits éditeurs Web qui n'ont pas les moyens d'assumer une grosse force de vente ou les coûts de recherche d'annonceurs. De leur côté, les annonceurs ont besoin des réseaux publicitaires pour promouvoir leurs produits efficacement auprès de publics pertinents, sans avoir à supporter les coûts importants de recherche d'éditeurs et de négociation directe avec chacun.

Pour mettre en relation les annonceurs avec les utilisateurs de contenus et services Internet, les réseaux publicitaires recourent à des stratégies contextuelles, verticales et comportementales. Les réseaux contextuels permettent aux annonceurs d'enchérir sur les mots clés figurant sur les sites Web des éditeurs de leur inventaire ; les réseaux verticaux regroupent les éditeurs de leur inventaire qui appartiennent au même domaine et les proposent aux annonceurs (les constructeurs automobiles voudront par exemple faire de la publicité sur des sites fréquentés par des amateurs de voitures) ; enfin, les réseaux comportementalistes recourent au ciblage comportemental pour adresser des publicités bien précises à certains utilisateurs en recueillant les données relatives à leur comportement de navigation sur plusieurs sites Web et en les utilisant pour établir un profil type du consommateur à cibler.

Les réseaux publicitaires sont opérés par différents types d'acteurs. Au départ, cette fonction était souvent assurée par un service spécialisé au sein des grosses agences de publicité (comme le groupe WPP, Omnicom, Publicis, etc.) ou par les grands portails Internet (Yahoo!, MSN, etc.). Depuis le milieu des années 2000, attirés par le développement de la publicité comportementale et de son potentiel, des réseaux publicitaires spécialisés ont fait leur apparition, et de grands acteurs Internet ayant accès à d'abondantes informations sur les utilisateurs se sont adjoint une activité de réseau publicitaire, souvent par croissance externe (par exemple, AOL via Advertising.com et TACODA ; Google via DoubleClick ; Yahoo! via Right Media et Blue Lithium, etc.). De plus en plus, les réseaux publicitaires de grande taille regroupent des moteurs de recherche, des médias et des fournisseurs de matériel.

Ces réseaux s'appuient sur des « plateformes ad-exchange » spécialisées dans la vente d'espaces et dans la vente de données, qui fonctionnent suivant un système d'enchères : les annonceurs enchérissent pour placer des publicités dans les espaces disponibles sur les sites web éditoriaux. Il existe aussi des bourses d'échanges de données, les « plateformes data-exchange » sur lesquels les annonceurs enchérissent pour avoir accès à des données clients. Ces données peuvent provenir du suivi des activités en ligne des utilisateurs et/ou d'une source non Internet (statistiques nationales, données de recensement, etc.). Elles sont de plus en plus souvent analysées et combinées, et un profil d'utilisateur est établi par des analystes spécialisés.

Dans un article du Wall Street Journal, Angwin et McGinty (2010) décrivent ce qu'ils appellent « l'écosystème du profilage » : lorsqu'un utilisateur visite un site Web, de petits fichiers de suivi (les « cookies ») observent ce qu'il fait et établissent le profil de son comportement. Les acteurs du profilage vendent ces informations, soit directement aux annonceurs, soit de plus en plus sur une plate-forme « data exchange », qui va pouvoir associer ces données à d'autres informations personnelles provenant de sources non Internet (données de recensement, registres fonciers, numéros d'immatriculation de véhicule, etc.). Ces données améliorées – qui prennent souvent la forme d'un « profil » d'utilisateur – sont ensuite revendues à des annonceurs qui recherchent des consommateurs correspondant au profil. En créant un lien entre le profil et le code d'identification unique contenu dans les cookies de l'ordinateur ou de l'appareil portable de l'utilisateur, l'annonceur peut servir à l'individu des publicités ciblées, et/ou acheter simplement de l'espace publicitaire sur des sites Web correspondant aux centres d'intérêt du profil de l'utilisateur via une plate-forme « ad-exchange ». Le procédé permet ainsi de présenter aux utilisateurs des publicités susceptibles de les intéresser.

Le ciblage comportemental est intéressant pour les annonceurs car il produit davantage de « hits ». D'après une étude de Beales (2010) sur 9 des 15 principaux réseaux publicitaires, la publicité comportementale représentait environ 18 % du total du chiffre d'affaires de la publicité en 2009 (595 millions USD), coûtait 2.68 fois

plus cher que la publicité aveugle et était plus de deux fois plus efficace à convertir les clics en achats effectifs – soit un taux de conversion de 6.8 % contre 2.8 % pour la publicité aveugle). Cette forme de publicité fournit en outre des informations sur les produits et services susceptibles d'intéresser le consommateur ciblé, réduisant ainsi les recherches et les coûts associés.

Les systèmes d'exploitation des smartphones constituent une plateforme relativement nouvelle de collecte de données et de publicité ciblée. Deux systèmes d'exploitation dominent : IOS d'Apple et Android de Google. Leurs créateurs respectifs ont également des intérêts commerciaux dans la publicité ciblée (encadré 2). Une activité aujourd'hui en pleine expansion consiste à regrouper les données recueillies à partir de téléphones portables et de smartphones et à établir le profil des utilisateurs. Mobclix, plate-forme d'achat et de vente d'espaces publicitaires, organise la rencontre de plus de 25 régies publicitaires avec quelque 15 000 applis cherchant des annonceurs. De plus, en localisant les téléphones, Mobclix devine plus ou moins où vivent leurs utilisateurs. Il lui suffit alors d'associer ces informations de localisation avec les données sociodémographiques et de dépenses recueillies hors ligne par la société Nielsen (Thurm et Yukari, 2010). Ce procédé permet d'accroître l'efficacité de la publicité mobile et l'utilité des informations pour les utilisateurs car elles sont en rapport avec le lieu où ils se trouvent.

Encadré 2. Plateformes pour smartphones

Avec le succès croissant et rapide des appareils mobiles connectés, la pertinence géographique des informations et des publicités est devenue l'un des grands axes d'innovation. Les deux plateformes les plus répandues sur les nouveaux smartphones sont iOS et Android, et leurs éditeurs, respectivement Apple et Google, possèdent les deux plus gros services – en chiffre d'affaires – d'envoi de publicité sur les téléphones portables.

Lors du test réalisé par le Wall Street Journal, c'est Google qui est arrivé en tête en nombre de données reçues par les applications des smartphones (Thurm et Yukari, 2010). Ses divisions AdMob, AdSense, Analytics et DoubleClick ont reçu des informations provenant de 38 des 101 applications testées. Google, dont les services publicitaires fonctionnent à la fois sur iPhone et sur Android, assure que ces entités conservent les données reçues séparément. Le principal réseau publicitaire de Google pour les appareils mobiles est AdMob, rachetée pour 750 millions USD en 2010. Avec AdMob, les annonceurs peuvent cibler les utilisateurs de téléphones en fonction de leur localisation, du type d'appareil utilisé et de leurs caractéristiques socioéconomiques, dont le sexe et la tranche d'âge.

La régie publicitaire iAd d'Apple ne fonctionne quant à elle que sur l'iPhone. Sur les 51 applications iPhone testées, 18 ont envoyé des informations à Apple. Apple envoie des publicités ciblées aux utilisateurs d'IPhone en fonction de ce qu'elle connaît d'eux grâce à son App Store et à son service musical iTunes. Les critères de sélection utilisés sont notamment les types de musique, de vidéos et d'applis téléchargées par l'utilisateur.

La distribution des informations traitées prend de nombreuses formes. S'agissant des principaux acteurs, une quantité importante des données recueillies et traitées est utilisée en interne par le logiciel de gestion des relations clients (CRM), qui s'appuie sur les programmes de fidélité et sur les historiques de transactions, de manière à proposer aux clients des offres de produits et de services adaptées. D'autres acteurs vendent et transfèrent les données qu'ils recueillent, ou encore achètent des données traitées sous forme de service (par exemple la cote de solvabilité, le profil, etc.) auprès d'analystes et de courtiers en données et/ou sur les places de marché de données (les plateformes « data exchange », voir encadré 3).

Dans le contexte de la publicité en ligne, une **plate-forme « data exchange »** est une place de marché sur laquelle les annonceurs enchérissent pour avoir accès à des données sur les consommateurs.[3] Ces données peuvent avoir été collectées au moyen du suivi des activités en ligne des utilisateurs ou à partir de sources non Internet (statistiques nationales, données de recensement, etc.). Par exemple dans le cas du pistage par un acteur tiers, la première fois qu'un site Web est visité par un utilisateur, le collecteur de données installe un cookie, qui attribue à l'ordinateur un numéro d'identification unique. Lorsque l'utilisateur visite un autre site affilié, le « pisteur » peut voir quels sites l'utilisateur a visités dans l'intervalle et donc établir un profil plus complet de ses habitudes et de ses goûts. Dans les secondes qui

suivent la visite d'un site Web, des informations détaillées sur l'activité de l'internaute peuvent être mises aux enchères sur une plate-forme « data exchange » de type BlueKai (Angwin, 2010). BlueKai, si l'on en croit son site web, est la première data exchange mondiale, avec des données relatives à plus de 300 millions d'utilisateurs, à travers plus de 30 000 attributs de données ; cette plateforme traite plus de 750 millions d'événements de données et enregistre plus de 75 millions d'enchères d'informations personnelles par jour.[4]

Encadré 3. Les courtiers en données

Les courtiers en données sont des entreprises qui rassemblent et fusionnent des informations individuelles agrégées, lesquelles sont ensuite vendues pour diverses utilisations : vérification du parcours professionnel d'un candidat, octroi d'un crédit ou exécution de décisions de justice, par exemple. Les données proviennent généralement de fichiers publics ou de sources variées auxquelles les individus ont apporté les informations. Dans la mesure où l'activité économique de ces entreprises repose généralement sur la vente de données à caractère personnel, le prix qui est attribué aux données par ces entreprises et par leurs clients dans la transaction peut fournir une approche de la valeur des données personnelles.

Le profil des courtiers en données est variable : il peut s'agir de prestataires de services spécialisés à destination des entreprises, de courtiers en informations, ou de simples services de localisation. Les prestataires de services spécialisés à destination des entreprises (comme LexisNexis) proposent des vérifications complètes et approfondies de toutes les informations dont peuvent avoir besoin les entreprises sur leurs partenaires commerciaux potentiels. Leur champ d'action va bien au-delà des seules données personnelles. Les courtiers en données (tels que Intelius ou Locate Plus) proposent aux consommateurs et aux petites entreprises différentes formules de fourniture d'informations à partir de fichiers publics et des données accessibles publiquement. Leurs produits permettent de retrouver des personnes, de vérifier l'identité de personnes rencontrées, de contrôler certains risques, de se protéger, etc. Enfin, les services de localisation (par exemple, LocatePeople.org, MelissaData.com et 123people.com) communiquent les adresses personnelles de personnes à des marchands de données, ou proposent de simples services permettant simplement de localiser des personnes, de connaître leur numéro de téléphone et leur adresse électronique, etc.

Le nombre d'entreprises proposant ces services semble en augmentation, mais il est difficile d'avancer des statistiques fiables car il n'existe pas de classification normalisée des courtiers en données. Certains se classent dans la catégorie professionnelle des « services d'hébergement et activités connexes », et d'autres dans celle des « agences d'évaluation de crédit ».

Au moment de la rédaction du présent rapport, l'organisation privacyrights.org recensait aux États-Unis 180 courtiers d'informations en ligne. Cela inclut des sociétés telles que LexisNexis, qui dit effectuer plus de 12 millions de vérifications par an, Acxiom, dont le chiffre d'affaires se monte à quelque 1.2 milliard USD par an, Experian, spécialisée dans les informations relatives au crédit, et beaucoup d'autres sociétés de plus petite taille. Il existe également des entreprises comme Everify.com, qui pour 19.95 USD fournissent des vérifications instantanées d'informations très précises telles que : nom et surnoms, date de naissance, historique de l'adresse sur une durée maximale de 40 ans, coordonnées téléphoniques actuelles et antérieures, associés, membres de la famille, voisins, casier judiciaire, contraventions, crimes, agressions sexuelles, procès, décisions rendues, patrimoine, inscriptions aux registres des mariages et des divorces, adresses électroniques et profils sur les réseaux sociaux. Le service Accurint de LexisNexis fournit quant à lui une liste détaillée des tarifs appliqués pour chaque produit, où l'on constate qu'un grand nombre de données personnelles isolées sont vendus à un prix dérisoire.

Usages

Une fois collectées, stockées et analysées, les données sont souvent mises à la disposition des utilisateurs finals, dernier maillon de la chaîne de valeur. Ceux-ci achètent généralement des profils d'individus (ou d'entreprises) pour compléter leurs propres activités commerciales. En évaluant le montant de ces transactions, on peut obtenir une approche de la valeur accordée aux données personnelles par le marché.

Les usages des données à caractère personnel par les entreprises, les organismes publics et les utilisateurs finals sont multiples. Bien que la classification de ces différents usages ne rentre pas dans le cadre du présent rapport, on peut néanmoins noter un certain nombre de grandes tendances. Le plus souvent, les données personnelles servent à améliorer la satisfaction des clients et l'efficience des transactions entre l'entité et les utilisateurs finals. Dans quelques secteurs tels que la santé, les transports et la sécurité, elles sont aussi agrégées pour accroître l'efficience opérationnelle et mettre en évidence des tendances macro.

Les *entreprises* utilisent généralement les informations pour mieux comprendre leurs clients et souvent pour leur servir des publicités ou des services ciblés. Les modèles d'entreprises sur Internet ont connu ces dernières années une mutation : au modèle de gratuité exposé par Chris Anderson dans son ouvrage *Free : entrez dans l'économie du gratuit* (Anderson, 2009) succède un modèle de non-gratuité décrit dans *Not-for-free : Revenue Strategies for a new world* (Berman, 2011). Le premier modèle s'appuie en général sur la publicité (encadré 1). Dans le second, Berman explique que les secteurs des médias et du divertissement ainsi que les éditeurs de contenus ont besoin de modèles d'entreprise plus diversifiés que l'approche couramment utilisée du « gratuit en ligne/payant hors ligne », ou que le financement assuré purement et simplement par la publicité. Berman *et al*. (2010) ont constaté que les modèles reposant sur la publicité en ligne ont une rentabilité unitaire moindre pour les éditeurs de contenus traditionnels : le lecteur d'un journal papier rapporte 18 fois plus que le lecteur du journal en ligne, et le spectateur d'une émission télévisée trois fois plus que celui visionnant un programme en ligne. Les auteurs expliquent cet écart financier par la pression à la baisse exercée sur les prix – en raison du trop grand nombre de contenus en ligne et du recours aux régies publicitaires –, et par le fait que les vidéos en ligne se prêtent beaucoup moins à la coupure publicitaire que les émissions de télévision (Berman *et al.*, 2010).

L'une des solutions proposées est d'adopter un modèle centré sur le consommateur, en utilisant les données dont on dispose pour lui offrir davantage de valeur. On va mettre en place des incitations pour obtenir du consommateur des informations le concernant, et lui offrir quelque chose d'intéressant en échange (Berman *et al.*, 2010). Il ressort de cette étude que les entreprises des médias et du divertissement ont besoin de trouver des méthodes novatrices et de moderniser les modèles économiques traditionnels – fondés sur le parrainage –pour attirer le consommateur de l'ère numérique. Il est notamment important qu'elles abandonnent la méthode du « multipoint » utilisée par les médias traditionnels et qu'elles tirent parti des avantages des plateformes numériques pour cibler les consommateurs et interagir avec eux de façon différenciée ou individualisée. Il leur faut passer à l'ère de « l'après-publicité » et privilégier les techniques de marketing centrées sur le consommateur.

Les *administrations et organismes publics :* Les pouvoirs publics utilisent de plus en plus de données personnelles – provenant de tierces parties mais aussi directement des individus concernés – pour administrer différents programmes. C'est le cas par exemple des services sociaux, des services fiscaux ou la délivrance de permis. Les données personnelles sont également couramment utilisées par les administrations pour une multitude d'opérations courantes : inscription sur les listes électorales, divulgation de l'identité des donateurs d'une campagne politique, vérifications d'identité des employés, exécution de l'obligation de versement d'une pension alimentaire. Autres utilisations possibles : pour la tenue des registres d'état civil et l'enregistrement des grands événements de la vie (naissance, mariage, divorce, adoption et décès), la gestion d'équipements collectifs tels que les voies à péage et les parcs nationaux. Les données à caractère personnel permettent de servir une population plus nombreuse, d'atténuer l'impression d'inégalité qui résulte d'évaluations subjectives, de réduire les coûts de l'exécution des décisions judiciaires et du maintien du niveau de qualification du personnel, ainsi que d'améliorer la traçabilité des décisions (Cate, 2008). La nouveauté réside dans l'utilisation de plus en plus généralisée de l'analyse et du rapprochement de données dans un objectif de coercition : on voit donc apparaître de nouveaux modes d'utilisation des données personnelles.[5]

Plusieurs des utilisations décrites précédemment présentent un intérêt pour les *utilisateurs finals*. Les modèles centrés sur le client utilisent les données personnelles et le ciblage comportemental pour améliorer la qualité des services fournis. La publicité ciblée réduit les coûts de recherche et de transaction en aidant les utilisateurs à trouver plus facilement les produits et les services qui les intéressent et leur épargne des publicités qui ne présentent aucun intérêt pour eux. Le fait que la publicité soit plus rentable est aussi un moyen de mieux financer des services gratuits (par exemple des services de recherche comme Bing), qui servent à de nombreux utilisateurs et qui sont financés par la publicité. Les réseaux sociaux sont eux aussi utiles pour les utilisateurs finals, non seulement dans leur vie privée mais aussi dans leur vie

professionnelle (par exemple, LinkedIn), car ils peuvent créer des opportunités : offres de collaboration et des débouchés professionnels, mais aussi vérifications par les employeurs éventuels sur les candidats à un poste.

Certains réseaux sociaux thématiques peuvent aussi rendre d'immenses services aux utilisateurs finals. PatientsLikeMe (encadré 4) en est un exemple : outre qu'il permet aux personnes souffrant d'une pathologie donnée de communiquer avec d'autres personnes dans la même situation, d'échanger avec elles et de recevoir du soutien, ce réseau fournit une base factuelle constituée de données personnelles qui peuvent être analysées, ainsi qu'une plateforme permettant de mettre en relation des malades avec des essais cliniques. Le principe consiste à faire coïncider les intérêts des malades avec ceux des professionnels : PatientsLikeMe vend à ses partenaires (notamment des sociétés pharmaceutiques et des fabricants d'équipements médicaux) des données agrégées anonymisées, afin de les aider à mieux comprendre le vécu des patients et les aspects concrets de l'évolution de la maladie. En outre, PatientsLikeMe partage les données relatives aux malades avec des chercheurs du monde entier. Le cas de PatientsLikeMe montre combien il peut être difficile de mesurer la valeur pécuniaire des données personnelles. Certes, on peut connaître les prix pratiqués pour la vente de données aux partenaires, mais pour les utilisateurs finals, le service est gratuit, alors que leurs échanges avec les autres malades ont beaucoup de valeur à leurs yeux. Or, cette valeur est extrêmement difficile à observer et à évaluer dans la mesure où les services sont fournis gratuitement.

Encadré 4. Un exemple réussi d'utilisation des dossiers médicaux personnels

Le réseau PatientsLikeMe est un exemple d'application/de plateforme Internet qui sert la recherche dans le domaine de la santé. Il s'agit une communauté Internet qui réunit des malades souffrant de pathologies lourdes, des médecins et d'autres parties prenantes. Ses membres échangent des informations qui peuvent être utilisées pour les recherches sur certaines maladies et ainsi améliorer la vie des malades. Cette plate-forme permet en particulier la collecte et le partage de données de terrain sur les malades et leur évolution. Les informations pertinentes sont transmises aux médecins, laboratoires pharmaceutiques, fabricants d'équipements médicaux, services de recherche et autres partenaires qui ont conclu un accord d'échange de données avec PatientsLikeMe. Dans le contexte de la médecine moderne, où la recherche a souvent besoin de grandes quantités de données, le réseau PatientsLikeMe apparaît comme un outil très précieux pour les professionnels de la santé et les différents acteurs de la lutte contre certaines maladies et bénéficie par-dessus tout aux malades eux-mêmes.

Source : www.patientslikeme.com/

Méthodes de détermination de la valeur des données à caractère personnel

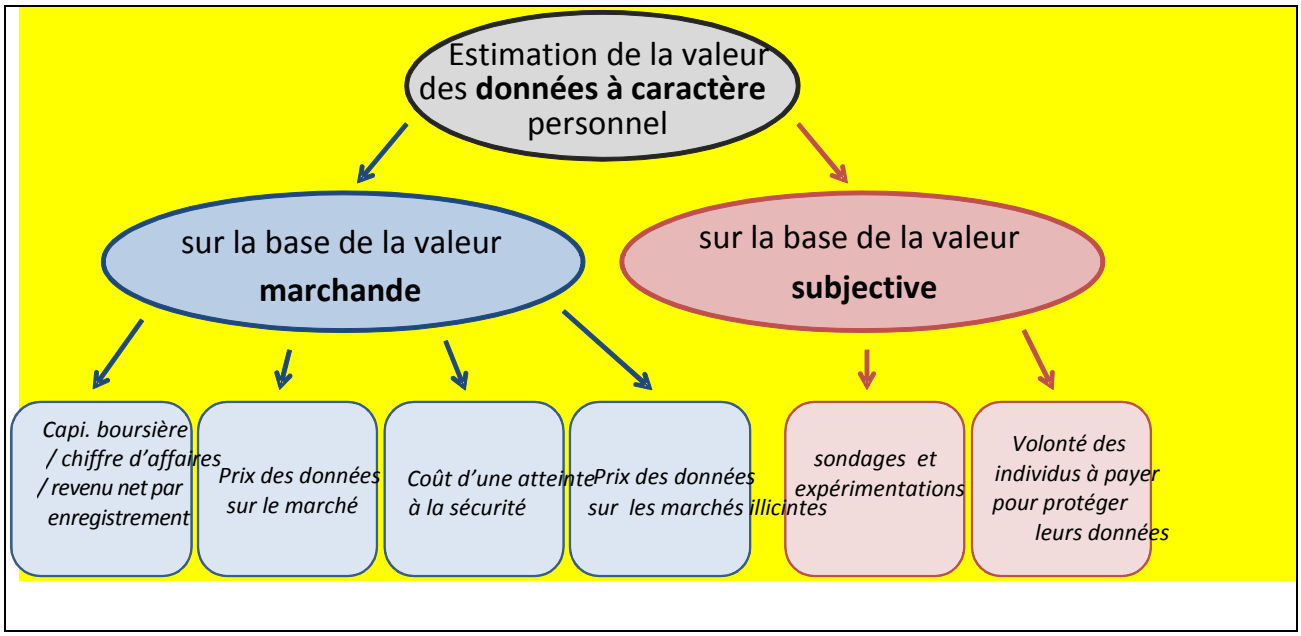
La présente section décrit les différentes méthodes que l'on peut utiliser pour déterminer la valeur pécuniaire des données personnelles. Nous l'avons vu, ces techniques ne mesurent pas les avantages sociaux et économiques des données personnelles, mais le prix qui leur est attribué sur le marché dans différents contextes. Cette section passe en revue les six méthodes pouvant permettre de réaliser une première estimation de la valeur pécuniaire des données, puis décrit les avantages et les inconvénients de chacune.

La plupart de ces méthodes s'intéressent à la valeur unitaire attribuée à un enregistrement ou à un utilisateur. Un enregistrement peut renfermer une seule information (par exemple l'âge de la personne), mais il peut aussi, dans un autre contexte, contenir un profil tout entier avec par exemple des informations sur les caractéristiques socioéconomiques et le parcours éducatif de la personne. Il n'est pas possible à ce stade de donner une définition, fût-ce-t-elle rudimentaire, de ce que contient un enregistrement. En revanche, les méthodes d'analyse examinent la valeur de chaque enregistrement dans le seul contexte du marché considéré.

Les méthodes d'évaluation sont rudimentaires et chacune mesure un aspect différent de la valeur pécuniaire ; elles ne sont donc pas directement comparables. Les transactions financières et les valorisations par le marché peuvent néanmoins servir de base pour appréhender plus largement l'importance des données personnelles dans l'économie.

Aucune méthode de calcul de la valeur des données personnelles ne fait l'unanimité. Ces méthodes utilisent : *i*) soit la valorisation par le marché des données personnelles ou d'autres mesures basées sur le marché ; *ii*) soit la valeur subjective attribuée par les individus à leurs données personnelles et à leur confidentialité (voir graphique 2).

Graphique 2. Estimation de la valeur des données à caractère personnel



Source : OCDE

Les mesures s'appuyant sur la valorisation **par le marché** sont des mesures de la valeur des enregistrements qui peut être relevée ou déduite de l'observation d'un marché. Comme proxies de la valeur des données personnelles on peut utiliser : *i*) un quotient de type capitalisation boursière par enregistrement, ou chiffre d'affaires par enregistrement, ou revenu net par enregistrement ; *ii*) le prix des données sur le marché ; *iii*) le coût d'une atteinte à la sécurité des données ; ou enfin *iv*) le prix des données sur les marchés illicites.

Les mesures axées sur une valorisation **par les individus** peuvent s'obtenir notamment au moyen de sondages et d'expérimentations économiques, ou en se fondant sur le consentement à payer des individus pour la protection de leurs données.

Deux points doivent être présents à l'esprit à cet égard. D'abord, il n'existe pas de mesure unique parfaite de la valeur des données personnelles. Chacune des mesures que nous citons peut être affectée par des biais dus à la méthode utilisée ou par des erreurs de mesure. Rappelons également qu'il est primordial de ne pas isoler les données personnelles du contexte dans lequel elles s'inscrivent, c'est-à-dire le modèle économique.

Deuxièmement, la plupart des statistiques dont on dispose sur les données personnelles concernent le marché des États-Unis. Par conséquent, dans la plupart des cas, les exemples fournis à titre d'illustration des mesures présentées ne fournissent que des estimations valables pour les États-Unis. Soulignons que toute extrapolation de ces exemples à d'autres marchés est délicate et devra tenir compte de facteurs géographiques importants, à commencer par les différences entre législations nationales en matière de vie privée.

Le tableau 1 fournit une vue synthétique des mesures précitées.

Tableau 1. Synthèse des mesures de la valeur des données à caractère personnel

Indicateur	Description	Avantages	Inconvénients possibles
Indicateurs fondés sur la valeur marchande			
Données financières par enregistrement de données	Capitalisation boursière (ou chiffre d'affaires, ou résultat net) de l'entreprise, divisée par le nombre total d'enregistrements de données personnelles utilisés par cette entreprise.	- Relativement facile à déterminer. - Traduit la réelle valeur économique générée par les données personnelles.	- Risques d'inexactitude car beaucoup d'autres facteurs peuvent avoir un impact sur la capitalisation boursière, le chiffre d'affaires ou le résultat d'une entreprise. - D'éventuels effets de synergie peuvent entraîner des surestimations dans le cas des entreprises disposant de grandes quantités de données. La pertinence de cette méthode dépend de la proportion du chiffre d'affaires directement liée aux données personnelles.
Prix des données sur le marché	Prix pratiqué sur le marché par les courtiers en données pour chaque entrée de données.	- Relativement facile à déterminer. - Traduit la valeur d'une entrée de données particulière sur le marché.	- Outre la valeur des données elles-mêmes, cette mesure inclut le coût de leur recherche et de leur traitement. Par ailleurs, elle ne tient pas compte du contexte dans lequel elles sont vendues, qui a une forte influence sur la demande (et le prix) des données.
Coût d'une atteinte à la sécurité des données	Coût économique d'une atteinte à la sécurité des données (pour les entreprises comme pour les individus) pour chaque entrée de données.	- Traduit une valeur de marché effective et une partie des risques contre lesquels les entreprises doivent se prémunir.	- Mesure le coût du préjudice causé par l'atteinte à la sécurité des données, plus que la valeur des données elles-mêmes. Cette mesure ne tient pas compte du préjudice subi par l'entreprise en termes d'image.
Prix des données sur les marchés illicites	Estimation du prix des données personnelles (par entrée de données) sur les marchés illicites.	- Traduit une valeur de marché d'une entrée de données particulière.	- Mesure difficile à établir et qui ne vaut que dans le contexte où les données sont réutilisées pour en tirer profit illicitement. Parce que les malfaiteurs doivent trouver un juste milieu entre cette valeur et le risque d'être repérés et condamnés, la valeur des données personnelles obtenue avec cette méthode risque d'être sous-estimée.
Indicateurs fondés sur la valorisation subjective par les individus auxquels se rapportent les données			
Sondages et expérimentations	La valeur pécuniaire des données personnelles est communiquée/révélée par les individus lors de sondages/d'expérimentations	- Aucune ambiguïté dans l'identification des données. - Reflète la valeur économique pure des données personnelles du point de vue des individus.	- Valeur hypothétique non validée par le marché. Des recherches montrent que l'évaluation par un individu des données personnelles le concernant est fortement dépendante du contexte : la formulation

	économiques.	- Les résultats peuvent généralement être utilisés dans le cadre d'études comparatives (entre économies et entre types de données).	des questions peut avoir une forte incidence sur les réponses.
Consentement à payer des individus pour protéger leurs données	Somme que les individus sont prêts à payer pour protéger leurs données personnelles.	- Reflète la valeur économique pure de la confidentialité du point de vue des individus.	- Mesure le coût du préjudice causé par une atteinte à la sécurité des données tel que perçu par les individus, plus que la valeur des données elles-mêmes.

Résultats financiers par enregistrement

La première méthode de calcul de la valeur des données personnelles consiste à utiliser les résultats financiers d'une entreprise (tels que capitalisation boursière, chiffre d'affaires et résultat net par utilisateur ou par enregistrement) pour déterminer la valeur marchande des données contenues dans chaque enregistrement. Ce type d'analyse n'est valable que pour les entreprises qui tirent la plupart – voire la totalité – de leurs revenus des données personnelles, ou pour celles qui comptabilisent séparément les recettes provenant d'activités liées aux données personnelles. Beaucoup d'entreprises qui travaillent avec des données personnelles ne sont pas cotées en bourse ou ne diffusent pas ces informations de façon suffisamment détaillée pour pouvoir procéder à ce type de mesure.

Chacun de ces indicateurs financiers mesure un aspect différent. La capitalisation boursière correspond à l'évaluation approximative de la valeur de l'entreprise sur le marché (établie sur la base du cours de ses actions), rapportée au nombre d'enregistrements qu'elle détient. Théoriquement, cette capitalisation reflète la valeur de l'entreprise à laquelle s'ajoute l'ensemble de ses bénéfices prévus ramenés à leur valeur actuelle. En réalité, il est fréquent que la valeur boursière d'une entreprise fluctue en fonction de l'humeur générale du marché et de perturbations économiques qui n'ont pas forcément de rapport avec la valeur intrinsèque des données personnelles.

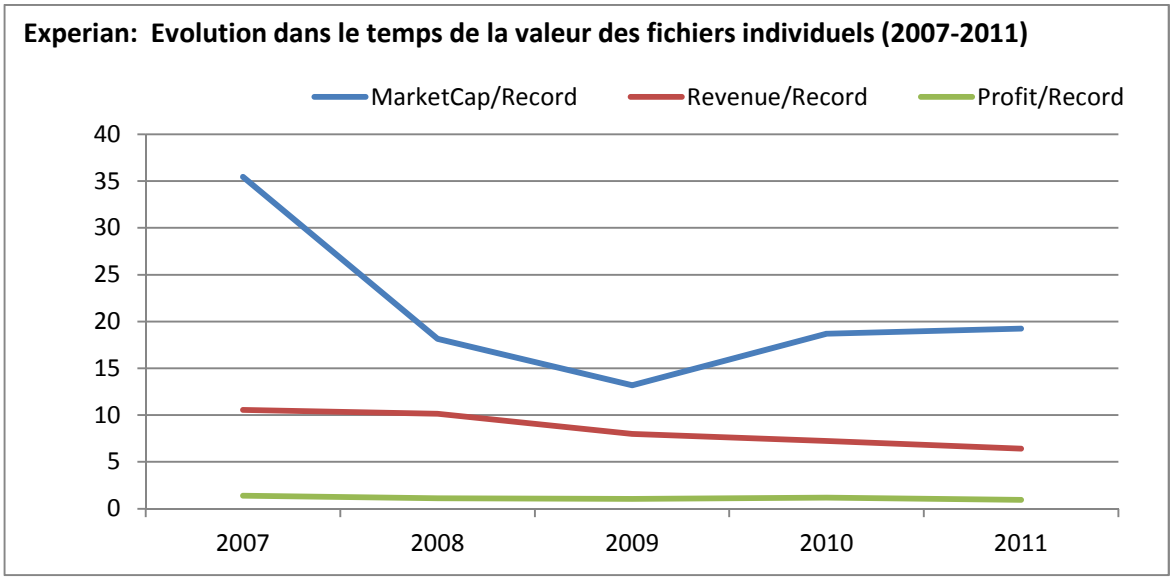
Le deuxième indicateur concerne les recettes déclarées de l'entreprise, rapportées au nombre total d'enregistrements. Cela équivaut à mesurer la « productivité » d'un enregistrement, autrement dit la somme d'argent qu'il rapporte chaque année. Cette méthode est parfois considérée comme fournissant une mesure plus réaliste de la valeur pécuniaire des données que la capitalisation boursière : elle rend compte en effet du montant moyen des recettes pouvant être attribué à chaque enregistrement de données, et est moins susceptible de subir les variations de l'humeur du marché et les perturbations générales de l'économie.

Le dernier indicateur correspond au montant des bénéfices par enregistrement de données et donc au bénéfice net (les recettes moins les coûts) enregistré pour chaque enregistrement détenu par une entreprise. De tous les indicateurs tirés des informations financières, les recettes par enregistrement seraient la source d'information la plus fiable car elles indiquent le montant (total) des sommes gagnées par l'entreprise grâce aux données, alors que les bénéfices par enregistrement incluent des coûts qui sont fluctuants et qui n'ont pas forcément de rapport avec la valeur intrinsèque des données personnelles.

À titre d'exemple, le courtier en données Experian a déclaré pour 2011 un chiffre d'affaires de 4.2 milliards USD et sa capitalisation boursière a oscillé entre 10 et 12 milliards USD, pour un nombre de dossiers – retraçant l'historique des demandes de crédit et des paiements – s'élevant à 600 millions pour les individus et 60 millions pour les entreprises. Cela nous donne une capitalisation de quelque 19 USD par enregistrement et un résultat annuel par enregistrement d'environ 6 USD. Le bénéfice par enregistrement était nettement moins élevé en 2001, de l'ordre de 1 USD par enregistrement (graphique 3).

Graphique 3. Experian : principales données financières par enregistrement

Dossiers relatifs aux individus et aux entreprises, millions (2007-2011)



Source : Rapports annuels d'Experian 2007-2011

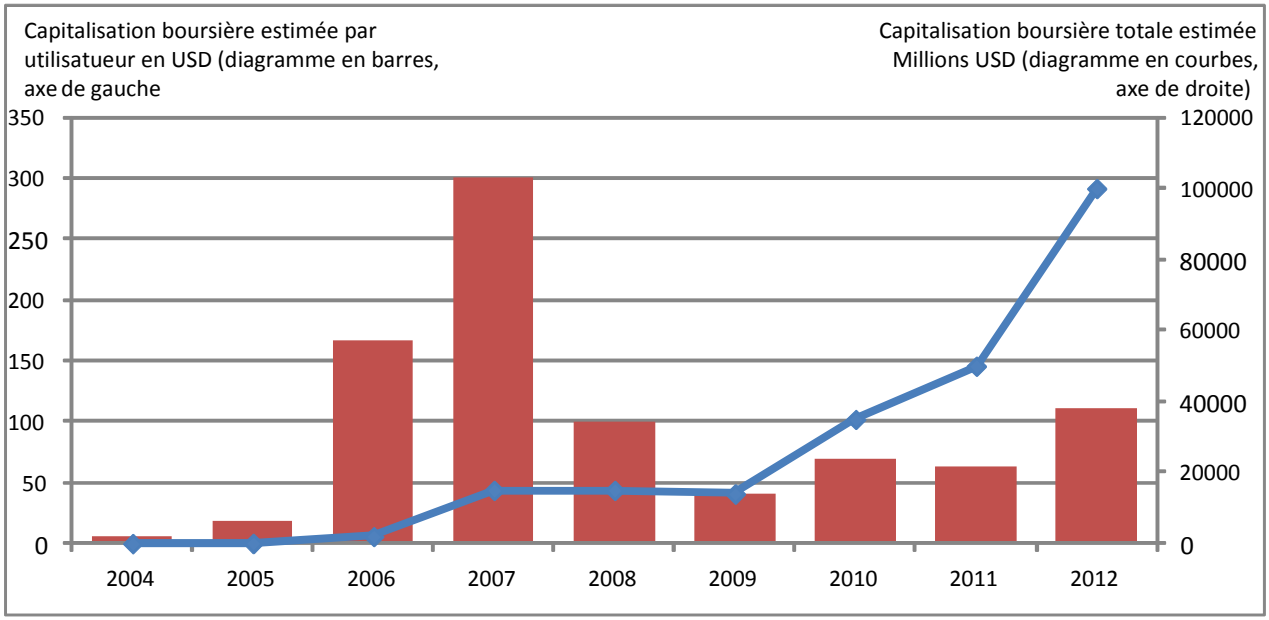
L'observation du graphique 3 nous montre que la capitalisation boursière par enregistrement d'Experian a fluctué de la même manière que l'évolution globale du marché entre 2007 et 2011. Après une forte baisse au début de la crise économique cette capitalisation a atteint son niveau le plus bas en 2009, avant de repartir à la hausse en 2011. En revanche, le résultat par enregistrement a suivi une courbe descendante plus lente et progressive entre 2007 et 2011. Le résultat annuel par enregistrement, de 10.55 USD en 2007, n'était plus que de 6.42 USD à l'issue de la période. Le bénéfice par enregistrement a également diminué, mais de façon moins importante : d'une moyenne de 1.40 USD en 2007, il est redescendu à 0.94 USD en 2011.

Il peut être utile de comparer les indicateurs financiers par enregistrement d'un fournisseur de données comme Experian avec ceux d'une autre entreprise dont la quasi-totalité des revenus proviennent de la publicité liée aux données personnelles, c'est-à-dire Facebook. En mai 2012, Facebook a fait son entrée en bourse avec une valorisation d'un peu plus de 100 milliards USD pour quelque 900 millions d'utilisateurs, ce qui équivalait à une valorisation de 111 USD par utilisateur enregistré.

Le graphique 4 montre que la valeur individuelle de chaque enregistrement (utilisateur) a beaucoup fluctué entre 2004 et 2012, si l'on se base sur les estimations de la valorisation boursière de Facebook. Il est important de noter que du fait que l'entreprise n'est cotée en bourse que depuis peu, la majorité des valorisations sont basées sur la taille des investissements hors marché pour un certain pourcentage des fonds propres. En dépit de la forte volatilité, la valorisation par utilisateur s'est maintenue entre 40 et 300 USD depuis 2006, tandis que la capitalisation boursière de l'entreprise passait dans le même temps de 2 à 100 milliards USD.

Graphique 4. Estimation de la capitalisation boursière par utilisateur de Facebook

2004-2012

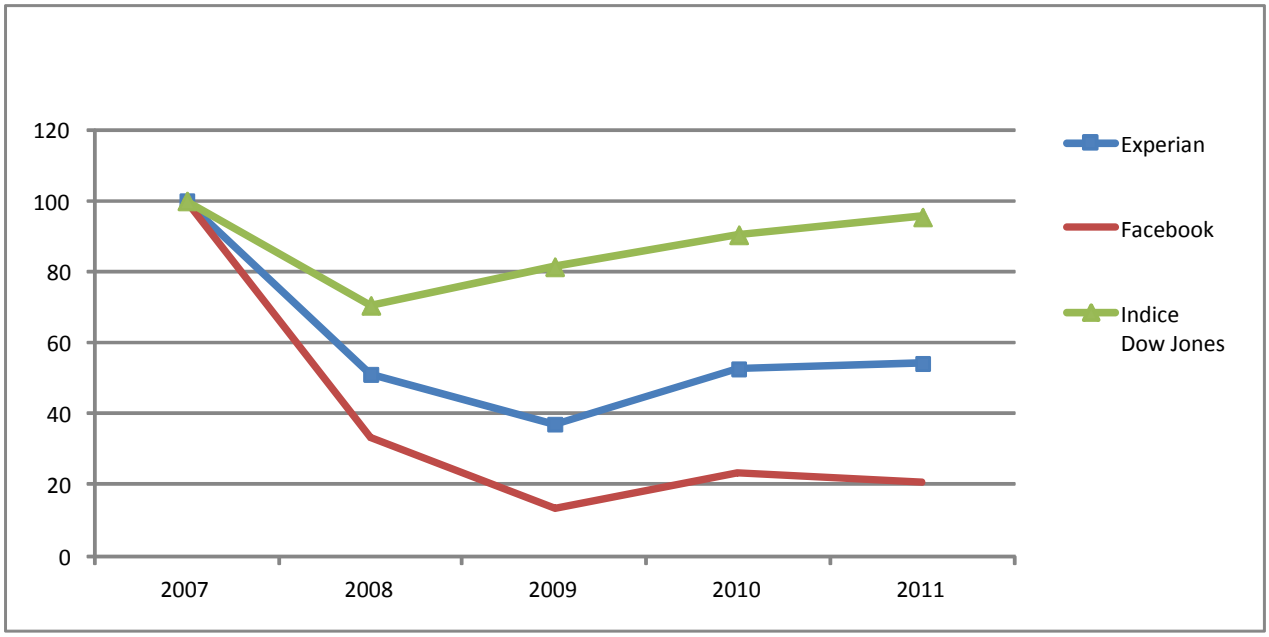


Source : Graphique élaboré par l'OCDE à partir des données de TechCrunch et CBS
http://techcrunch.com/2011/01/10/facebook-5/
http://www.cbsnews.com/8301-505250_162-57370133/number-of-active-users-at-facebook-over-the-years/

Le fournisseur de données Experian a connu une évolution similaire de la valeur par enregistrement entre 2007 et 2011. Le graphique 5 présente la valeur normalisée par enregistrement ou par utilisateur d'Experian et de Facebook, de manière à montrer leur évolution dans le temps par rapport à l'indice Dow Jones, qui reflète la tendance générale du marché boursier aux États-Unis. S'agissant de la valorisation par enregistrement ou par utilisateur, Facebook et Experian suivent une courbe similaire mais marquent un retard par rapport à la reprise générale affichée aux États-Unis par l'indice Dow Jones.

Graphique 5. Évolution dans le temps de la capitalisation boursière par enregistrement/utilisateur ainsi que de l'indice Dow Jones

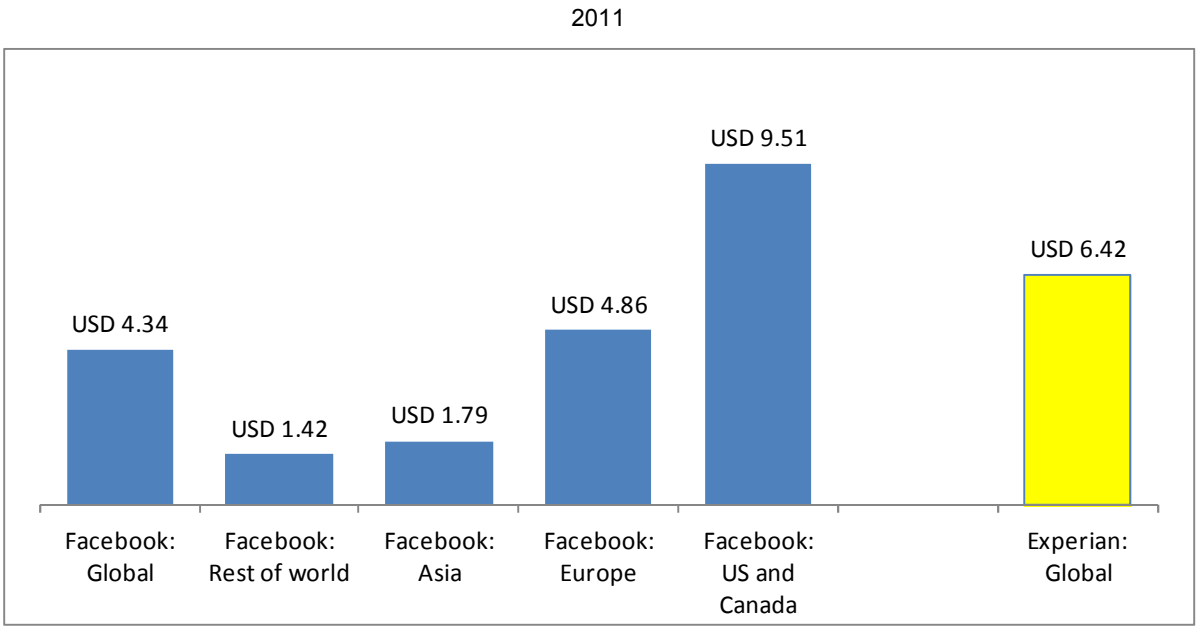
Indice : 2007 = 100



Les chiffres de la capitalisation boursière d'Experian et de celle attribuée à Facebook semblent confirmer la thèse selon laquelle l'évaluation des données personnelles provenant de la capitalisation boursière est à la merci de l'humeur générale du marché et des perturbations extérieures qui n'ont pas forcément de rapport avec la valeur intrinsèque des données.

Le résultat par enregistrement ou par utilisateur fournit une meilleure idée de la valeur monétaire des données car elles sont directement liées à la somme payée par un tiers pour accéder aux données. Experian et Facebook utilisent des modèles économiques très différents, ne stockent pas les mêmes types de données et utilisent différemment les données recueillies. Il est donc intéressant de voir que, malgré ces différences, ces deux entreprises ont enregistré des niveaux comparables de résultat par utilisateur (ou par enregistrement) (graphique 6).

Graphique 6. Résultat annuel par utilisateur/enregistrement, Facebook et Experian



Note : Les chiffres de Facebook correspondent au revenu publicitaire par utilisateur, et proviennent de TechCrunch : http://techcrunch.com/2012/05/03/stats-facebook-made-9-51-in-ad-revenue-per-user-last-year-in-the-u-s-and-canada/.
Les chiffres d'Experian proviennent du rapport annuel 2011 de l'entreprise.

Avantages de cette méthode

Étant donné que les entreprises cotées en bourse communiquent leurs états financiers et que les investissements de capital-risque sont souvent rendus publics, il est relativement facile de mesurer la valeur des données à partir des résultats financiers. Les obligations d'information auxquelles sont soumises les entreprises publiques impliquent que des données sur la capitalisation boursière, sur les recettes et le résultat net sont disponibles pour un grand nombre d'entreprises. Ces données peuvent être associées au volume des données personnelles utilisées par une entreprise particulière pour calculer à titre d'exemple la valeur sur le marché d'un seul enregistrement de données personnelles.

Un autre avantage de cette méthode est que les résultats financiers sont produits dans un contexte marchand, ce qui signifie que la mesure obtenue traduit la valeur ajoutée économique générée par l'utilisation effective des données personnelles sur ce marché. Autrement dit, le bénéfice net, le chiffre d'affaires ou la capitalisation boursière sont les résultantes des forces du marché. Par conséquent, une mesure de la valeur des données personnelles qui s'appuie sur ces résultats financiers peut être considérée comme un dérivé de la véritable valeur marchande des données.

Inconvénients de cette méthode

Il existe plusieurs restrictions à l'utilisation des résultats financiers (par exemple la capitalisation boursière, le chiffre d'affaires ou le bénéfice net) – divisés par le nombre d'utilisateurs ou d'enregistrements – pour mesurer la valeur des données personnelles. De nombreux éléments autres que les données personnelles ont une influence sur les résultats financiers d'une entreprise, comme par exemple les ressources humaines et le capital physique, la quantité d'actifs immatériels et l'expertise. L'utilisation des

résultats financiers peut par conséquent entraîner une importante surestimation de la valeur des données, en particulier dans les secteurs où les données à caractère personnel ne sont pas le principal facteur de production.

Le choix du type de mesure financière a son importance dans le calcul de la valeur des données personnelles. L'utilisation du chiffre d'affaires par donnée personnelle/utilisateur présente bel et bien des limites car le chiffre d'affaires ne contribue à la croissance que dans la mesure où il génère de la valeur ajoutée (ou un surplus). Le résultat net par donnée personnelle/utilisateur peut fournir une mesure plus satisfaisante car il rend compte de façon plus précise de la valeur ajoutée créée par l'entreprise en utilisant les données. L'utilisation de la capitalisation boursière par donnée personnelle est peut-être la technique qui donne la mesure la moins fiable de la valeur des données personnelles, car pour la majorité des entreprises, cette capitalisation est brouillée par les nombreux autres éléments qui interviennent dans la chaîne de valeur de chaque entreprise.

Autre inconvénient, il n'existe pas forcément un rapport linéaire entre le volume de données personnelles et les résultats financiers, même pour les entreprises qui utilisent beaucoup ce type de données. Pour citer un exemple, l'utilisation des données personnelles dans l'activité économique peut produire des synergies : ainsi, dans une entreprise donnée, la valeur économique d'un enregistrement de données isolé sera moins élevée que celle du même enregistrement au sein d'un ensemble homogène d'enregistrements. Ces synergies risquent à leur tour de fausser considérablement les estimations.

Enfin, les résultats financiers d'une entreprise peuvent aussi être très influencés par les tendances du marché, les perturbations aléatoires et la spéculation. Cela signifie que la mesure de la valeur des données personnelles peut être imprécise et fluctuer au fil du temps en fonction des tendances générales du marché ou de l'activité spéculative, sans que cela n'ait de rapport avec la valeur intrinsèque des données.

Prix des données sur le marché

L'un des principaux moyens de comprendre la valeur pécuniaire des données est d'observer leur prix sur les marchés concurrentiels. Le prix des données sur le marché est déterminé par la rencontre entre l'offre et la demande. Or, les données personnelles sont un bien économique « non divisible » ou « non substituable », ce qui veut dire que leur utilisation par une personne ne diminue pas le stock de biens ; un même enregistrement peut être vendu un grand nombre de fois à un grand nombre de clients, et il peut être utilisé plusieurs fois par le même client. Par conséquent, le prix sur le marché d'un enregistrement vendu à un client ne traduit pas la valeur pécuniaire à part entière des données qui le composent, mais donne en fait une idée du prix d'équilibre payé par chaque client pour une copie des données. La recette moyenne par enregistrement que l'on obtient avec la première méthode de calcul est comparable à la somme des prix qui ont été payés pour un enregistrement par l'ensemble des clients pendant une année.

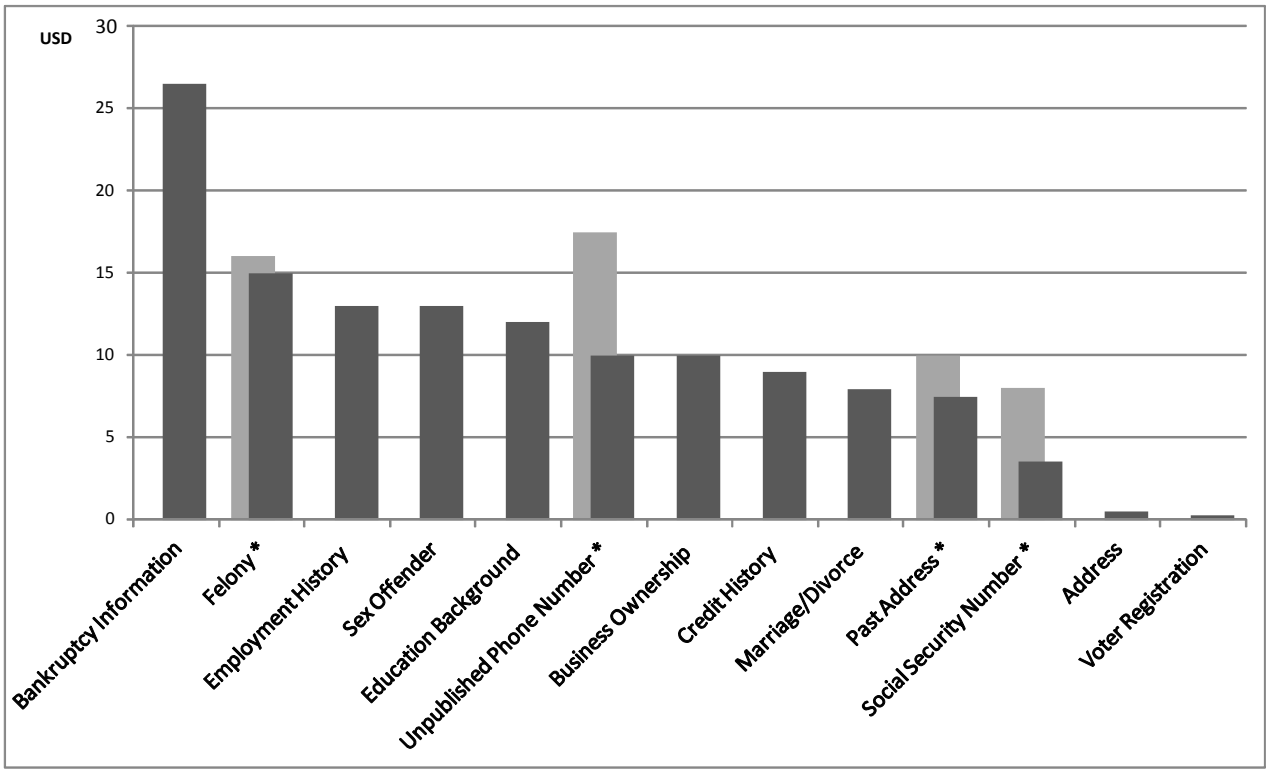
Autrefois, les transactions de données personnelles s'effectuaient généralement de gré à gré entre entreprises ; aujourd'hui, en revanche, Internet ouvre de nouveaux marchés et permet de se renseigner plus facilement sur les prix. Pour certains types de données personnelles, les prix sont consultables sur Internet (c'est le cas par exemple pour les vérifications des antécédents personnels), alors que pour d'autres, il faut s'entretenir avec un vendeur et obtenir un devis pour le type d'information souhaité. Un exemple représentatif est celui du marché américain, où de nombreuses entreprises – les courtiers en données – se lancent dans cette activité et vendent des données personnelles. [6]

L'une des particularités des courtiers en données est qu'ils réalisent des transactions en utilisant des données provenant d'entreprises tierces. Les prix pratiqués par ces entreprises fournissent donc une indication intéressante de la valeur marchande d'un enregistrement. Le graphique 7 reproduit des estimations provenant de diverses bases de données en ligne (par exemple, Aristotle, LexisNexis,

DocuSearch, Experian, Merlin Data et Pallorium). A l'heure de la rédaction du présent rapport, les données personnelles suivantes étaient proposées aux prix suivants : 0.50 USD pour une adresse, 2 USD pour une date de naissance, 8 USD pour un numéro de sécurité sociale (numéro d'identification national), 3 USD pour un numéro de permis de conduire et 35 USD pour un livret militaire. Le prix groupé d'une adresse, d'une date de naissance, d'un numéro de sécurité sociale, d'un dossier de crédit et d'un livret militaire est estimé à 55 USD. Il s'agit uniquement des estimations, mais elles ont le mérite de donner une idée de la valeur marchande relative des différentes catégories de données personnelles.

Graphique 7. Prix sur le marché des différentes catégories de données personnelles

(Par enregistrement)



* Les fournisseurs pratiquent deux prix différents.

Sources : Locate Plus (adresse, numéro de téléphone sur liste rouge, infractions pénales), Pallorium (adresse, adresse passée, numéro de téléphone sur liste rouge, numéro de sécurité sociale), KnowX via Swipe Toolkit (adresse passée, situation maritale, situation financière, entreprises détenues), LexisNexis via Swipe Toolkit (études, antécédents professionnels, numéro de sécurité sociale, infractions pénales, agressions sexuelles), Experian (antécédents en matière de crédit) et Voters online.com (inscription sur les listes électorales).

Avantages de cette méthode

Le prix des enregistrements de données personnelles pratiqué sur le marché présente deux grands avantages qui en font un instrument utile pour mesurer la valeur économique des données personnelles.

Premièrement, à l'instar des résultats financiers par enregistrement, le prix des données sur le marché est relativement facile à interpréter lorsque les données sont disponibles. Plusieurs courtiers en données affichent publiquement le prix qu'ils pratiquent pour différents enregistrements de données personnelles. Il est ensuite facile de regrouper ces prix et de les comparer pour en déduire indirectement la valeur marchande d'un enregistrement.

Deuxièmement, comme beaucoup d'entreprises dans d'autres secteurs, les courtiers en données opèrent sur des marchés concurrentiels. Le prix qu'ils demandent pour les données personnelles est donc la résultante de la rencontre entre l'offre et la demande. Cela signifie qu'il traduit au moins partiellement le prix réel qui se pratique sur le marché pour se procurer une donnée particulière.

Inconvénients de cette méthode

Il existe un certain nombre d'inconvénients à l'utilisation des prix du marché – provenant de plusieurs sources – pour déterminer la valeur pécuniaire des données. Le prix du marché incorpore également les coûts (notamment salariaux) supportés par le fournisseur de données et éventuellement l'effort de recherche épargné à l'acheteur. Les données peuvent être disponibles à partir d'une source publique, mais elles nécessitent des efforts de traitement non négligeables pour être exploitables par les clients.

Le prix auquel se négocient les données sur un marché ouvert ne s'applique qu'à un contexte particulier. La valeur pécuniaire d'un numéro de téléphone peut, en soi, être relativement faible si l'on prend cette donnée isolément. En revanche, lorsque ce numéro est associé au niveau de revenu et à un ensemble de centres d'intérêt, il aura probablement plus de valeur pour certains acteurs du marché. Cela veut donc dire que les prix observés sur le marché doivent être replacés dans leur contexte.

Enfin, la qualité des données provenant des courtiers en données ne peut être vérifiée a priori et ces données risquent de manquer de justesse, de fiabilité ou de précision. Ce risque est souvent intégré dans le prix des données, ce qui fausse encore plus leur véritable valeur.

Marchés illicites

Une méthode assez proche de l'approche par le prix du marché que nous venons de décrire est celle qui consiste à déterminer la valeur des données personnelles sur les marchés illicites (cybercriminalité). Ces marchés fonctionnent comme des forums en ligne (généralement en utilisant les canaux de discussion des serveurs de messagerie instantanée de type IRC [Internet Relay Chat]) sur lesquels les cybermalfaiteurs achètent et vendent des biens et des services aussi variés que du code de maliciels, des ordinateurs zombies, des attaques de déni de service, l'hébergement de pages illicites, et surtout, des données personnelles telles que « numéros d'identification nationaux, numéros de cartes de crédit, comptes utilisateur, listes d'adresses électroniques et numéros de comptes bancaires » (voir Symantec, 2010 et Panda Security, 2010). Des sociétés de sécurité comme Panda Security (2010) ont repéré et observé ces marchés ainsi que les offres qui y sont faites.

Ces observations permettent de mettre en lumière la gamme des valeurs attribuées par les cybermalfaiteurs aux données personnelles. Selon Symantec (2010), par exemple, les informations relatives aux cartes de crédit étaient en 2009 les données les plus fréquemment mises en vente sur les marchés de la cybercriminalité, puisqu'elles représentaient 19 % de l'ensemble des données proposées (voir le tableau 2).[7] Les prix pratiqués étaient compris en 2009 entre 1 et 30 USD. À en croire Symantec, la valeur des données personnelles dépend de plusieurs facteurs, comme par exemple les caractéristiques socioéconomiques de l'individu et l'intensité de ses transactions en ligne, notamment pour ce qui est des achats et des opérations bancaires.

Plus important encore, le prix des informations relatives aux cartes de crédit varie dans le temps en fonction du nombre total d'offres disponibles sur les marchés de la cybercriminalité. Ce nombre dépend à son tour de la quantité de données personnelles ayant été volées lors d'atteintes à leur sécurité (voir l'encadré 5). Les informations relatives aux cartes de crédit sont parmi les données les plus volées et revendues. En 2008, elles représentaient environ un tiers des données proposées à la vente sur les marchés de la cybercriminalité, mais leur proportion varie d'année en année en fonction du nombre d'atteintes à la

sécurité des données commises. Selon Symantec (2010), il y a eu plus d'attaques de ce genre et donc d'informations relatives aux cartes de crédit en vente en 2008 qu'en 2007 ou 2009. Compte tenu de la baisse du nombre d'atteintes à la sécurité et de la quantité de données disponibles en 2009, le prix maximal demandé pour des informations relatives aux cartes de crédit a augmenté de 16 %, pour atteindre plus de 30 USD.

Tableau 2. Biens et services mis en vente sur les marchés de la cybercriminalité

Classement mondial en 2009	Type de donnée	2009	2008	Gamme de prix en USD
1	Numéros de cartes de crédit	19 %	32 %	0.85 - 30
2	Références de comptes bancaires	19 %	19 %	15 - 850
3	Comptes de messagerie électronique	7 %	5 %	1 - 20
4	Adresses de messagerie électronique	7 %	5 %	1.70/Mo - 15/Mo
5	Scripts Shell	6 %	3 %	2 - 5
6	Données d'identification complètes	5 %	4 %	0.70 - 20
7	Copies de cartes de crédit	5 %	2 %	4 - 150
8	Services d'envoi de messages électroniques	4 %	3 %	4 - 10
9	Services de retrait	4 %	3 %	0 - 600 plus 50-60 %
10	Codes d'administration de sites Web	4 %	3 %	2 - 30

Source : Symantec (2010)

L'une des grandes différences entre les marchés illicites de données personnelles comme les cartes de crédit, et les marchés licites est que les données obtenues illégalement peuvent être considérées comme des « biens rivaux » ; en effet, un numéro de carte de crédit aura moins de valeur pour les malfaiteurs si cette carte a déjà été utilisée par d'autres, ce qui accroît le risque de repérage et réduit les possibilités d'utilisation frauduleuse du numéro. Il en résulte que le prix d'équilibre appliqué pour les données illicites à un seul client est sans doute plus proche de la véritable valeur marchande des données que lorsqu'il s'agit de données licites pouvant être revendues plusieurs fois sans que cela ne diminue leur utilité pour d'autres clients.

Avantages de cette méthode

Le principal avantage de cette méthode est de fournir un prix reflétant la valeur marchande d'une donnée précise. Bien que le marché soit illicite – ce qui implique que les prix qui y sont pratiqués peuvent difficilement être comparés aux prix en vigueur sur les marchés licites (cf. les inconvénients) –, le prix relatif d'une donnée particulière ainsi que les facteurs intervenant dans son évaluation peuvent être riches d'enseignement. D'autre part, les prix pratiqués sur le marché noir peuvent refléter implicitement les externalités mieux que d'autres techniques d'évaluation.

Inconvénients de cette méthode

Un certain nombre d'inconvénients limitent l'utilité de cette méthode. Tout d'abord, les prix pratiqués sur les marchés illicites sont difficiles à recueillir. Compte tenu du caractère illicite des biens et des services proposés, les prix et les transactions ne sont jamais véritablement transparents et il n'est donc pas facile de procéder à une évaluation. Les estimations risquent fort par conséquent d'être faussées du fait de l'absence d'échantillons représentatifs. Par ailleurs, les prix ne sont valables que dans le contexte dans lequel les données sont réutilisées pour en tirer d'autres bénéfices illégalement. Parce que les malfaiteurs doivent prendre en compte le risque d'être repérés et condamnés, la valeur des données personnelles obtenue avec cette méthode peut être surestimée.

Encadré 5. Les atteintes à la sécurité des données et leurs impacts économiques pour les entreprises et les consommateurs

Lorsque des données personnelles sont recueillies, stockées ou traitées, les incidents de sécurité qui surviennent peuvent porter gravement préjudice à la vie privée – comme l'ont montré de récentes affaires très médiatisées[8] –, mais ils s'avèrent aussi coûteux pour les entreprises que pour les utilisateurs. Selon les informations dont dispose *Privacy Rights Clearinghouse*, le nombre d'incidents répertoriés se situe aux alentours de 45 par an, et le nombre total d'enregistrements volés est de plus en plus lié à des infractions de grande ampleur, c'est-à-dire touchant plus de 10 millions d'enregistrements.[9] Lorsque l'on tient compte des sommes d'argent qui sont associées aux procès en cours et autres mesures visant à réduire les préjudices directs et indirects, le coût par donnée volée peut fournir une mesure simplifiée du niveau de risque auquel sont confrontées les entreprises en fonction du nombre de données personnelles qu'elles stockent.

L'attaque qui a visé TJX, par exemple – avec quelque 100 millions d'enregistrements – a obligé la société à provisionner 118 millions USD sur l'exercice 2008 pour couvrir les coûts et dommages-intérêts possible (1.18 USD par enregistrement).[10] Outre les 11 millions USD (9 % du total) en honoraires de conseil en sécurité et autres dépenses liées aux attaques, cette somme incluait un fonds de secours de 107 millions USD (1.07 USD par enregistrement) destiné à couvrir les dommages-intérêts liés aux procès en cours (voir Leyden, 2007). Elle ne couvrait en revanche pas le préjudice en termes d'image, l'impact sur la marque, les coûts d'opportunité et autres conséquences indirectes. En revanche, TJX a annoncé l'impact qu'avaient eu les intrusions avec ses résultats du deuxième trimestre 2008, avec un bénéfice net en baisse de 57 % par rapport au même trimestre 2007, résultat qui s'explique sans doute par des charges plus élevées liées à cette atteinte.

Un autre cas intéressant est celui de l'atteinte à la sécurité des données dont a été victime Heartland Payment Systems en 2009, et qui a touché quelque 130 millions d'enregistrements. Suite à cette attaque, la société a décidé de constituer un fonds de 105 millions USD pour couvrir les dommages et intérêts (0.80 USD par enregistrement). Cette somme était répartie comme suit : 41 millions USD (39 %) pour les clients MasterCard, 60 millions (57 %) pour les clients VISA et presque 4 millions (4 %) pour les clients American Express (voir Leyden, 2010). On ne connaît pas le montant des investissements de Heartland Payment Systems en sécurité, ni celui de ses coûts indirects. En revanche, le rapport financier de l'exercice 2009 révèle que la société a accusé une perte nette de plus de 52 millions USD (contre un bénéfice net de 42 millions USD au cours de l'exercice 2008), malgré un chiffre d'affaires en hausse de 7 % par rapport à l'année précédente. La valeur de ses actions a également reculé, passant de 18 USD le 2 janvier 2009 à 15.44 USD le 16 janvier, puis à 8.54 USD le 23 janvier, soit une chute de presque 50 % en trois semaines (l'attaque a été annoncée par Heartland Payment Systems le 20 janvier 2009).

Un exemple plus récent est celui de l'atteinte à la sécurité dont ont été victimes le réseau PlayStation Network de Sony et Sony Online Entertainment en 2011, qui a exposé au grand jour 103 millions d'enregistrements et entraîné la fermeture du réseau PlayStation Network pendant 23 jours. Selon les responsables de Sony, cette infraction coûtera à la société un minimum de 171 millions USD (soit 1.7 USD par enregistrement). Il est intéressant de noter que cette somme ne couvre pas le paiement des dommages et intérêts comme dans les exemples précédents, mais plutôt les « dépenses liées à la mise en œuvre d'un programme de prévention de l'usurpation d'identité et les offres promotionnelles visant à reconquérir les clients, entre autres » (Goodin, 2011) ; en d'autres termes, cette enveloppe permettra de financer (en partie) les coûts indirects (préjudice en termes de réputation et coûts d'opportunité). Si l'on suppose que Sony doit en plus constituer un fonds correspondant à 1 USD par enregistrement en dommages-intérêts pour les procès en cours, cela veut dire que la société devra trouver 103 millions supplémentaires.[11] Cette somme n'inclura toujours pas les investissements dans les programmes d'évaluation et d'amélioration de la sécurité (par exemple les honoraires des conseillers en sécurité).

Les atteintes à la sécurité des données entraînent un coût non seulement pour les entreprises qui en sont victimes, mais aussi pour les consommateurs. Il a ainsi été indiqué que 10 % des Américains ont fait l'objet d'une usurpation d'identité et que chacun d'eux a perdu en moyenne 5 000 USD.[12] De la même manière, une étude réalisée récemment sur 2 500 Australiens a montré qu'une personne sur dix avait été victime d'un vol d'identité en ligne au cours des 12 derniers mois, et que chacune d'elles avait perdu en moyenne 1 000 AUD.[13] Au Royaume-Uni, quelque 2 millions de personnes font

l'objet chaque année d'une usurpation d'identité, pour un coût total au niveau national de 2.7 milliards GBP. Sachant que les malfaiteurs gagnent en moyenne 1 000 GBP en espèces ou en nature pour chaque nom dérobé, la somme précitée se compose pour l'essentiel de l'argent que rapporte aux malfaiteurs l'utilisation des identités usurpées, et pour le reste des frais engagés par les individus et les entreprises pour prévenir et détecter les infractions puis réparer les dégâts. Dans les cas graves, il faut parfois plus de 200 heures pour résoudre les problèmes occasionnés par une fraude à l'identité, soit l'équivalent d'une année de congés.

Sondages et expérimentations économiques

Autres approches de la valeur pécuniaire des données personnelles, l'expérimentation économique[14] et la réalisation de sondages, qui permettent de connaître le prix qu'une entreprise doit débourser pour qu'un individu lui communique des informations personnelles.

Ces dernières années, plusieurs études expérimentales ont été menées pour chiffrer la valeur attribuée par les individus à leurs données personnelles dans différents contextes. Ce domaine de recherche n'est pas encore très développé, car l'évaluation individuelle de la confidentialité est une notion très complexe et très dépendante du contexte, donc très difficile à analyser dans un cadre expérimental. Bien que la recherche dans ce domaine n'en soit encore qu'à ses débuts, elle peut déjà fournir deux informations d'ordre général :

- Premièrement, les individus ont tendance à faire la différence entre :
    - la *valeur qu'ils attachent à leurs données personnelles* (c'est-à-dire la somme d'argent qu'ils jugent suffisante pour révéler les données les concernant) ;
    - la *valeur qu'ils attachent à la confidentialité de ces données* (c'est-à-dire la somme d'argent qu'ils sont prêts à payer pour empêcher que leurs données personnelles ne soient divulguées).
- Deuxièmement, la détermination de ces deux valeurs est extrêmement sensible aux contextes et l'on ne peut en attendre une absolue certitude et précision.

La distinction entre la valeur des données personnelles et la valeur de la confidentialité a une grande importance sur le plan matériel. Concernant *la valeur de la confidentialité*, les individus ont bel et bien la possibilité de payer pour empêcher que leurs données personnelles ne soient divulguées. Les utilisateurs d'Internet peuvent par exemple acheter un logiciel permettant d'anonymiser leur navigation et de masquer leur comportement en ligne, même si cela se fait souvent au détriment de leur vitesse de navigation. Les individus ont également de nombreuses occasions de divulguer leurs informations personnelles en échange de compensations pécuniaires : c'est ce qui correspond à *la valeur des données personnelles*. Certaines sociétés offrent par exemple un lot de produits aux individus qui les autorisent à effectuer le pistabe de leurs activités sur Internet.

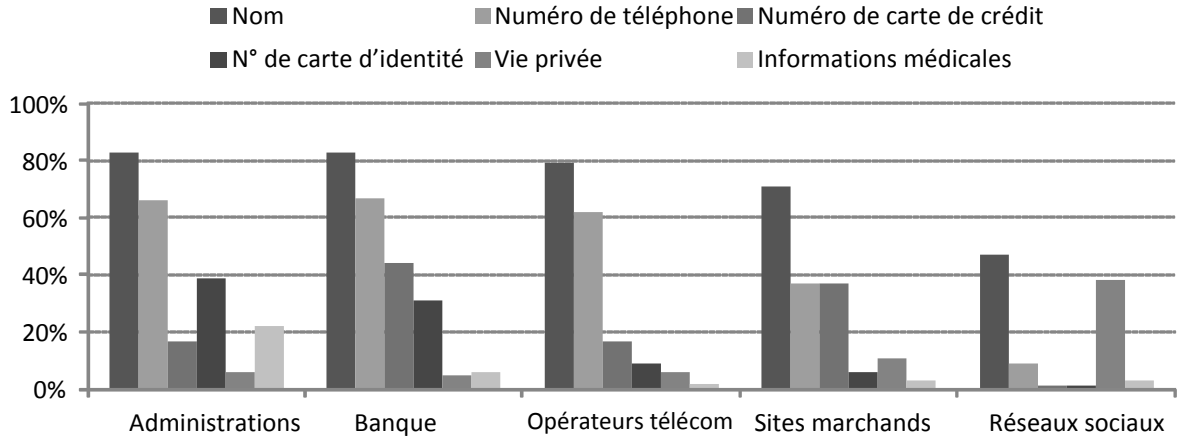
De nombreuses études empiriques portent à la fois sur *la valeur des données personnelles* et *la valeur de la confidentialité*. C'est le cas par exemple des travaux de Spiekermann *et al.* (2001), Chellappa et Sin (2005), Wathieu et Friedman (2005), Huberman *et al.* (2006), Cvrcek *et al.* (2006) et Hui *et al.* (2007). En règle générale, la proportion de consommateurs qui refusent de renoncer à la confidentialité de leurs données en échange d'une somme d'argent est plus grande que celle des consommateurs qui acceptent, pour le même montant, de payer pour la protection de cette confidentialité (Acquisti *et al*., 2009). Comme l'ont constaté Hui et Png (2006), la différence entre *la valeur des données personnelles* et *la valeur de la confidentialité* aux yeux des individus pourrait expliquer en partie les conclusions disparates de ces études empiriques.

Ces études montrent par ailleurs que la valeur *de la confidentialité* et la valeur *des données personnelles* sont extrêmement sensibles au contexte. Certains de ces travaux montrent que même les individus clairement soucieux la confidentialité de leurs données peuvent être convaincus de communiquer des informations sensibles les concernant à des personnes qu'ils ne connaissent pas (Spiekermann *et al.*, 2002). Citons à cet égard l'expérience de terrain réalisée par Infosec, au cours de laquelle 71 % des employés de bureau empruntant la station Liverpool Street à Londres étaient prêts à divulguer le mot de passe de leur adresse électronique en échange d'une barre chocolatée (Infosec Europe, 2004). L'application stricte du principe économique de la « révélation des préférences » pourrait conduire à la conclusion que les individus ne se soucient ni de la préservation de leur confidentialité, ni de leurs données personnelles (Rubin et Lenard, 2002).

Cette apparente incohérence pourrait s'expliquer à l'aide des outils de l'économie comportementale, qui tient compte du contexte et des motivations psychologiques des individus dans ce contexte. En utilisant ces techniques, plusieurs études comme celles de Schwarz (1999) et Acquisti *et al.* (2009) montrent à quel point *la valeur de la confidentialité* et *la valeur des données personnelles* peuvent être influencées par des éléments contextuels qui, on pourrait le croire, devraient jouer un rôle limité dans la prise de décision. Acquisti *et al.* (2009) mettent également en évidence la complexité des préférences individuelles en matière de confidentialité en étudiant la distribution des valeurs implicites de la confidentialité et des données elles-mêmes. Leur constat est que ces évaluations ne suivent pas une distribution normale ou uniforme, mais décrivent une courbe en U, concentrées autour des extrêmes et de certaines valeurs focales.

Les sondages montrent en outre que l'impact des incidents dépend du contexte, en particulier du type d'information affecté par l'incident et du secteur dans lequel la transaction en ligne a eu lieu. On voit également dans certains de ces sondages que les utilisateurs accordent moins d'importance à la divulgation en ligne de leur nom ou de leur numéro de téléphone, qu'à celle de leur numéro de carte de crédit, de leur numéro de carte d'identité ou d'informations sur leur vie privée. Les individus attachent une valeur particulièrement élevée aux informations médicales (graphique 8).

Graphique 8. Importance accordée par les individus aux données personnelles en France, par type d'information et par secteur



Source : Sondage CDC/IDate réalisé auprès de 700 utilisateurs d'Internet âgés de plus de 15 ans, octobre 2009.

Comme le montrent un ensemble d'expérimentations réalisées par Frog (2011) aux États-Unis, en Inde et en Chine, il existe trois catégories de données que les individus sont enclins à protéger, le niveau de protection (et la valeur attribuée) différant pour chacune d'elles : i) la première catégorie comprend les numéros de sécurité sociale (numéros de carte d'identité) et le numéro de carte de crédit, auxquels la plupart des individus accordent une valeur élevée (de 150 à 240 USD par donnée) ; ii) la deuxième catégorie inclut l'historique des opérations numériques, comme par exemple l'historique de navigation sur

le Web, ainsi que la localisation et les informations médicales (environ 50 USD) ; iii) la troisième et dernière catégorie correspond aux faits se rapportant aux individus, notamment l'historique des achats en ligne et celui des clics effectués sur des publicités en ligne, auxquels les individus accordent peu de valeur (de 3 à 6 USD).

La même étude révèle que toutes les marques et tous les secteurs ne bénéficent pas du même niveau de confiance. Le récipiendaire des données importe beaucoup. Les personnes interrogées lors du sondage réalisé par Frog (2011) ont déclaré qu'elles faisaient davantage confiance aux établissements financiers, aux entreprises de technologie, aux opérateurs de télécommunications mobiles et aux fabricants de mobiles. C'est aux administrations publiques et aux réseaux sociaux qu'elles font moins le confiance.

Avantages de cette méthode

Les résultats des sondages et des expérimentations économiques présentent plusieurs avantages en tant qu'instrument de mesure de la valeur des données personnelles pour les individus. Les expérimentations économiques et leurs procédures sont bien décrites dans la littérature (voir Kagel et Roth, 1997), et il est donc relativement facile de mettre en évidence et d'interpréter les évaluations qu'elles produisent. Les expériences sont conçues de façon à inciter les participants à révéler la valeur qu'ils attribuent aux données les concernant. Ce type de mesure doit donc être à même de refléter la valeur économique pure, non faussée, des données personnelles du point de vue des individus.

D'autre part, les résultats peuvent généralement être utilisés pour des études comparatives (entre pays et entre types de données). Le cadre neutre de l'expérimentation permet en effet de répéter une expérience donnée sur plusieurs échantillons de participants et d'utiliser les résultats à des fins de comparaison.

Inconvénients de cette méthode

L'un des principaux inconvénients de l'utilisation des sondages pour obtenir indirectement une valeur des données personnelles est qu'il n'y a pas validation par le marché. En fait, ces mesures correspondent à la valeur hypothétique et spéculative des données personnelles telle qu'elle a été communiquée ou déclarée par les individus. Elles fournissent non pas une valeur économique confirmée par les forces de l'offre et la demande, mais une évaluation possible résultant d'une vision subjective et individuelle. Les sondages et expérimentations ont également mis clairement en évidence que les valeurs attribuées aux données personnelles sont fortement dépendantes du contexte. Ainsi, le prix auquel les individus sont prêts à vendre les informations les concernant dépend de leur perception de la manière dont les données seront utilisées et de la confiance qu'ils ont dans l'entité qui les reçoit.

Consentement à payer révélé pour la protection des données (assurance)

Une autre méthode pour attribuer une valeur économique aux données personnelles d'un individu consiste à déterminer combien cette personne serait prête à payer pour protéger ces données. On peut pour cela observer les marchés proposant des polices d'assurance de protection contre l'usurpation d'identité. On remarque ainsi que le courtier en données Experian évoqué plus haut propose aux États-Unis un service de ce type – ProtectMyID –, facturé 155 USD pour une durée d'un an.[15].

D'autres sociétés proposent un service de suppression du nom des individus des bases de données marketing. C'est le cas de Reputation.com,[16] qui permet aux utilisateurs de retrouver et d'effacer des informations personnelles sur des sites Web dont ils ne veulent pas – ou plus – qu'ils possèdent leurs données. Ce service est facturé quelque 140 USD par an.

Une autre innovation récente est l'apparition de logiciels et de services anti-pistage, qui risquent d'obliger les professionnels du pistage et du ciblage à passer à un système d'autorisation pour pouvoir faire

leur travail. À ce jour, les dispositifs de refus de la publicité ciblée sont limités, soit parce qu'ils fonctionnent assez mal (par exemple, du fait qu'ils s'appuient sur l'utilisation des cookies, que les utilisateurs avertis doivent supprimer et/ou bloquer), soit parce qu'ils s'appliquent à la publicité ciblée et non au suivi à proprement parler (par exemple en proposant de mettre fin au ciblage, mais pas au suivi). Cela étant, un nombre croissant de sociétés proposent des logiciels et des services qui limitent ou empêchent le suivi. Il peut s'agir de modules relativement simples qui sont ajoutés au navigateur pour bloquer la publicité ou mettre en évidence et gérer les cookies (Internet et Flash). Ces modules additionnels sont également disponibles pour les utilisateurs de messagerie électronique. Certaines entreprises offrent des solutions de prévention : c'est le cas par exemple de SafetyWeb, qui propose un service de suivi parental des activités des mineurs sur les réseaux sociaux, au prix de 100 USD par an.[17]

Avantages de cette méthode

L'un des avantages évidents de l'utilisation de la volonté déclarée des individus de payer pour protéger leurs données pour obtenir indirectement la valeur des données personnelles est qu'elle reflète la valeur économique pure d'une violation de la confidentialité du point de vue des individus. La valeur déclarée par les individus ne subit l'influence ni des frais de transaction (recherche ou autres), ni des fluctuations du marché, mais traduit uniquement la perception individuelle de la valeur financière d'une éventuelle atteinte à la confidentialité.

Inconvénients de cette méthode

L'utilisation du consentement à payer ou « prix psychologique » de la protection des données pour les individus pour déterminer la valeur économique des données personnelles présente deux risques de distorsion.

De même que pour les sondages et les expérimentations économiques, le consentement à payer déclaré par les individus pour la protection de leurs données a l'inconvénient de ne pas être validé par le marché. Cette mesure traduit la volonté d'un individu de payer (*valorisation*), mais elle ne s'exprime pas sur un marché et n'est pas confirmée par des transactions commerciales (*valeur*). Ainsi, si une personne donnée indique un consentement à payer faible pour assurer sa confidentialité, il se peut qu'il n'y ait aucune compagnie d'assurance sur le marché qui propose cette garantie pour un prix suffisamment faible.

Deuxièmement, le consentement à payer révélé des individus pour protéger leurs données mesure davantage le coût (perçu par les individus) du préjudice causé par l'atteinte à la sécurité des données, que la valeur des données elles-mêmes. En vérité, un grand nombre d'études économiques insistent sur la différence entre *la valorisation des données personnelles* et *la valorisation de la confidentialité* (Acquisti *et al.*, 2009 ; Hui et Png, 2005). Ces deux concepts, bien que liés, sont différents; il convient donc d'être très prudent lorsque les mesures de *la valeur de la confidentialité* sont utilisées comme valeur de substitution à celle des données personnelles.

Enfin, la question se pose de l'efficacité réelle ses services censés prévenir et réparer le vol de données personnelles. Cela laisserait entendre que les consommateurs sont davantage prêts à payer moins cher pour un service aux résultats douteux que plus cher pour un service plus fiable.

Conclusions et suite des travaux

Le présent rapport est une première étape dans la compréhension des différentes méthodes permettant de mesurer la valeur pécuniaire des données personnelles. Il ne s'intéresse pas aux avantages sociaux et économiques qui peuvent découler de l'utilisation des données personnelles mais ne se reflètent pas dans les transactions commerciales. Il n'examine pas non plus d'autres indicateurs du marché des données personnelles, par exemple l'augmentation des investissements de capital-risque attirés par le potentiel

analytique de l'utilisation des données personnelles, ou par le développement de nouveaux produits ou services bénéficiant des gains d'efficience permis par l'analyse des données. Cette première étude aboutit toutefois à un certain nombre de constatations intéressantes qui pourraient justifier la poursuite du projet.

Nécessité d'améliorer les données et la collaboration

Nous manquons de données fiables pour réaliser l'analyse qui permettrait de mieux comprendre la valeur pécuniaire des données personnelles. Cette compréhension pourrait être utile à l'élaboration des politiques publiques. De surcroît, il peut être intéressant pour les pouvoirs publics comme pour les entreprises tout au long de la chaîne de valeur de travailler ensemble pour cerner la valeur potentielle des données personnelles.

Pour comprendre cette valeur potentielle, il est primordial de disposer de données de meilleure qualité. Les entreprises qui opèrent au long de la chaîne de valeur ont une bonne connaissance du nombre d'enregistrements qu'elles possèdent et ont besoin d'avoir une vision stratégique de la valeur économique de ces enregistrements. Il est par conséquent justifié que les gouvernants, les chercheurs et les entreprises engagent une collaboration pour mieux comprendre les avantages et les risques que peut présenter cette accumulation de données personnelles.

Compréhension des différents contextes régionaux

Pour obtenir un tableau plus complet de la valeur des données personnelles dans différents contextes régionaux, une analyse plus approfondie serait nécessaire. La plupart des statistiques dont on dispose pour le moment et qui servent d'illustration à l'évaluation des données personnelles se rapportent au marché américain. Toute extrapolation de ces exemples à d'autres marchés doit tenir compte des paramètres régionaux, c'est-à-dire non seulement des différences entre les régimes nationaux en matière de vie privée, mais aussi de facteurs socio-économiques comme le niveau d'information et d'éducation des individus, ou de l'intensité de la concurrence qui prévaut sur le marché.

Utilité des études de cas pour comprendre les effets

L'un des thèmes récurrents tout au long de ce rapport est le caractère contextuel de l'évaluation des données personnelles. La conséquence est que les effets macroéconomiques sont beaucoup plus difficiles à analyser du fait de l'absence de données harmonisées et de mesures d'impacts au fil du temps.

Au lieu des effets macroéconomiques, on peut partiellement étayer l'analyse sur des études portant sur des cas spécifiques afin d'observer les effets produits dans différents contextes, ce qui peut permettre d'extraire plus d'informations utiles au législateur. Ces études peuvent porter sur des secteurs particuliers, comme l'utilisation des données personnelles en recherche médicale. Elles peuvent également être associées à des ensembles de données spécifiques (par exemple les données relatives aux réseaux sociaux ou au parcours de navigation, utilisées à des fins particulières).

Une rentabilité probablement non linéaire, ce qui dénote des effets de réseau

Il est vraisemblable que la valeur pécuniaire, économique et sociale des données personnelles augmente de manière non linéaire par rapport au nombre des enregistrements. La valeur d'un enregistrement isolé peut être très faible, mais la valeur et l'utilité de cet enregistrement vont en s'accroissant à mesure qu'augmente le nombre d'enregistrements avec lesquels il peut être comparé. Ces effets de réseau ont des conséquences pour les politiques publiques, car on peut tirer beaucoup plus de valeur du même enregistrement extrait d'une grande série que d'une petite. Cela peut avoir des répercussions en matière de concurrence et sur d'autres aspects importants comme la portabilité des données.

Un surplus de l'entreprise et du consommateur difficile à quantifier

Même si l'on réussit à calculer la valeur pécuniaire des données personnelles, celle-ci ne traduira pas tous les avantages sociaux et économiques que peuvent procurer ces données. En effet, le surplus du consommateur – c'est-à-dire la différence entre ce qu'un utilisateur est prêt à payer pour un bien ou un service et le prix du marché – ne se reflète pas dans les prix sur les marchés concurrentiels. Cela veut dire qu'une analyse qui ne prendrait pas en compte le surplus du consommateur sous-estimerait probablement les véritables avantages sociaux et économiques.

Le surplus du producteur se reflétera dans le bilan des entreprises qui utilisent les données, mais là aussi, il pourra être difficile à isoler des données d'exploitation générales de l'entreprise. L'examen d'études de cas bien précises et une coopération plus étroite avec les professionnels des données pourraient compléter notre compréhension des impacts des données personnelles.

Des marchés dans lesquels les individus peuvent contrôler et vendre leurs propres données se développent : ils deviendront des sources d'informations

1. D'autres nouveautés qui se profilent pourraient procurer de nouvelles valorisations des données personnelles et nous renseigner sur leur valeur de marché. Les *data lockers (*ou « coffres de données ») proposés par certaines entreprises sont des dispositifs qui permettent aux utilisateurs de stocker en ligne (et de modifier à leur guise) les données qu'ils souhaitent partager avec des tiers en échange d'un pourcentage du produit de la vente de leurs données. Ces coffres de données offrent une meilleure transparence quant à la collecte, la vente et l'utilisation des données. Les utilisateurs pourraient être disposés à partager *encore plus* de données personnelles s'ils considèrent qu'ils peuvent ainsi avoir un meilleur contrôle de la manière dont elles seront utilisées et s'ils retirent un avantage social ou économique évident en le faisant. Il s'agit d'une nouvelle piste, et on ignore encore si ces coffres de données pourront être un modèle économique viable, mais c'est une piste à explorer.

RÉFÉRENCES

Acquisti, A., L. John et G. Loewenstein (2011) What is Privacy Worth?, document interne disponible à l'adresse : http://pages.stern.nyu.edu/~bakos/wise/papers/wise2009-6a1_paper.pdf, consulté le 18 mai 2012.

Anderson, C. (2009) Free: The future of a radical price, Hyperion. Disponible à l'adresse : http://www.longtail.com/the_long_tail/2009/07/free-for-free-first-ebook-and-audiobook-versions-released.html (novembre 2010).

Angwin, J. (2010) The Web's New Gold Mine: Your Secrets, Wall Street Journal, 30 juillet 2010. Disponible à l'adresse : http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html (janvier 2011).

Angwin, J. et McGinty, T. (2010) « Sites Feed Personal Details To New Tracking Industry », The Wall street Journal, 30 juillet 2012, disponible à l'adresse : http://online.wsj.com/article/SB10001424052748703977004575393173432219064.html

Angwin, J. et Steel, E. (2011) « Web's hot new commodity: Privacy », The Wall Street Journal, 28 février 2011. Disponible à l'adresse : http://online.wsj.com/article/SB10001424052748703529004576160764037920274.html?mod=WSJ_Tech_MIDDLETopNews (septembre 2010).

Appel, J.M. (2008), "Why shared medical database is wrong prescription", Orlando Sentinel, December 30, 2008 http://en.wikipedia.org/wiki/Electronic_health_record

Beales, H. (2010), « The Value of Behavioral Targeting », Network Advertising Initiative, disponible à l'adresse : http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf

Berman, S.J. (2011) Not for Free: Revenue strategies for a new world, Harvard Business School Press.

Berman, S.J., Battino, B. et Feldman, K. (2010) Beyond content: Capitalizing on the new revenue opportunities, IBM Institute for Business Value, Disponible à l'adresse : http://ibm.com/iibv

Boyd, D. (2010) « Privacy, Publicity, and Visibility », Microsoft Tech Fest. Redmond, 4 mars. Disponible à l'adresse : http://www.danah.org/papers/talks/2010/TechFest2010.html (octobre 2010).

Burridge, N. (2010) Annual cost of identity theft is £2.7bn, The Independent, 18 octobre 2010. Disponible à l'adresse : http://www.independent.co.uk/news/uk/crime/annual-cost-of-identity-theft-is-16327bn-2109431.html (avril 2011).

Cate, F.H. (2008) Government Data Mining: The Need for a Legal Framework, in Harvard Civil Rights-Civil Liberties Law Review, Vol. 43. Disponible à l'adresse : http://harvardcrcl.org/ (septembre 2011).

Chellapa, R. et R.G. Sin, 2005. « Personalization Versus Privacy: An Empirical Examination of the Online Consumers' Dilemma », Information Technology and Management, 6(2-3), pp. 181-202.

Cvrcek, D., M. Kumpost, V. Matyas et G. Danezis, 2006. « A Study On The Value Of Location Privacy », compte rendu d'un atelier sur la confidentialité dans la société électronique (WPES '06), pp. 109-118.

CyberSource (2011) 2011 Online Fraud Report, CyberSource. Disponible à l'adresse : www.pymnts.com/cybersource-2011-online-fraud-report-12th-annual-edition-online-payment-fraud-trends-merchant-practices-and-benchmarks (avril 2011).

FEM (2011) Personal Data: The Emergence of a New Asset Class, Forum économique mondial. Disponible à l'adresse : http://www.weforum.org/issues/rethinking-personal-data (mai 2011).

FTC (2009) Self-Regulatory Principles for Online Bhavioral Advertising, FTC, Washington DC. Disponible à l'adresse : http://www.ftc.gov/opa/2009/02/behavad.shtm (septembre 2010).

Frog (2011), The Value of Personal Data: A Global Perspective, Presentation at Mobile World Congress, février 2011.

Goodin, D. (2011), « PlayStation Network breach will cost Sony $171m », The Register, 24 mai 2011, disponible à l'adresse http://www.theregister.co.uk/2011/05/24/sony_playstation_breach_costs/

Hui, K.-L., H-H. Teo, S.-Y. Lee, 2007. « The Value of Privacy Assurance: An Exploratory Field Experiment », MIS Quarterly, 31(1), pp. 19-33.

Hui, K-L. et Png, I.P.L. (2006) The Economics of Privacy, in Hendershott, T. (dir. pub.) Handbooks of Information Systems, Volume 1. Elsevier.

Infosoc (2004), « Passwords for a Chocolate Bar », compte rendu d'étude disponible à l'adresse : http://www.net-security.org/secworld.php?id=2075

Kagel, J. H. et A. E. Roth (1997), « The Handbook of Experimental Economics », Princeton University Press.

Leyden, J. (2010), « Heartland coughs $41m to settle MasterCard claims », The Register, 20 mai 2010 disponible à l'adresse http://www.theregister.co.uk/2010/05/20/heartland_mastercard_settlement/

Narayanan A. et V. Shmatikov (2010) Privacy and security Myths and fallacies of "Personally identifiable information," Communications of the ACM, Vol. 53 (6).

Neate, R. et Mason, R. (2009), Networking site cashes in on friends, The Telegraph, 31 janvier 2009. Disponible à l'adresse : http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/4413483/Networking-site-cashes-in-on-friends.html (avril 2011).

OCDE (2010), The role of Internet Intermediaries in Advancing Public Policy Objectives, atelier, Paris, juin 2010, OCDE DSTI/ICCP(2010)13.

Ohm, P. (2010), Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review, 57.

PWC (2010) IAB Internet Advertising Revenue Report: 2010 First Half-Year Results, the Interactive Advertising Bureau, octobre 2010. Disponible à l'adresse : http://www.iab.net/insights_research/947883/adrevenuereport (octobre 2010).

Schneier (2010), SecuritySchneier on Security, A blog covering security and security technology. Disponible à l'adresse : http://www.schneier.com/blog/archives/2009/11/a_taxonomy_of_s.html

Schwarz, N. (1999), « Self-reports: How the questions shape the answers », *American Psychologist*, 54(2).

Sinclair, L. (2010) « Facebook targets $2bn as it overtakes Google », The Australian, 22 mars 2010. Disponible à l'adresse : http://www.theaustralian.com.au/business/media/facebook-targets-2bn-as-it-overtakes-google/story-e6frg996-1225843478239 (septembre 2010).

Spiekerman, S., Grossklags, J. and Berendt, B. (2002), « E-privacy in 2nd generation E-Commerce », ACM Conference on Electronic Commerce (EC'01), Hrsg. ACM New York, 38-47. Tampa, Florida: ACM Press.

Steel, E. et Angwin, J.(2010) The Web's Cutting Edge, Anonymity in Name Only, Wall Street Journal, 4 août 2010. Disponible à l'adresse : http://online.wsj.com/article/SB10001424052748703294904575385532109190198.html (novembre 2010).

Symantec (2010), « The Silent Epidemic: Cybercrime Strikes More Than Two-Thirds of Internet Users », communiqué de presse, 8 septembre. Disponible à l'adresse : www.symantec.com/about/news/release/article.jsp?prid=20100908_01

Symantec (2011), « Symantec Internet Security Threat Report: Trends for 2010 », Volume 16, avril.

Thurm, S. et Yukari, I.K. (2010) Your Apps Are Watching You, Wall Street Journal, 17 décembre 2010. Disponible à l'adresse : http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html?mod=what_they_know (janvier 2010).

Tucker, C. (2010) The Economic Value of Online Customer Data, OCDE, Paris. Disponible à l'adresse : http://www.oecd.org/document/22/0,3746,en_2649_34255_46565782_1_1_1_1,00.html (janvier 2011).

Tynan, D. (2007) As Applications Blossom, Facebook Is Open for Business, Wired Magazine, juillet 2007. Disponible à l'adresse : http://www.wired.com/techbiz/startups/news/2007/07/facebook_platform (novembre 2011).

Wathieu, L. et A. Friedman, 2005. « An Empirical Approach to Understanding Privacy Valuation », compte rendu du quatrième atelier sur la dimension économique de la sécurité de l'information (WEIS '05).

FEM (2011) Personal Data: The Emergence of a New Asset Class, Forum économique mondial. Disponible à l'adresse : http://www.weforum.org/issues/rethinking-personal-data (mai 2011).

NOTES

[1] Le Cahier des charges de la révision des lignes directrices de l'OCDE régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel (2011) est disponible ici : http://www.oecd.org/dataoecd/63/29/48975226.pdf

[2] La dernière étape du cycle de vie des données personnelles est leur élimination, aspect fondamental des mesures de protection de la confidentialité et de la sécurité des informations. Elle évoquée dans cette étude et ne fait pas l'objet d'un examen à part entière car elle n'apporte pas de contributions évidentes à la chaîne de valeur.

[3] What They Know: A Glossary, *Wall Street Journal*, 31 juillet 2010. Disponible à l'adresse : http://online.wsj.com/article/SB10001424052748703999304575399492916963232.html (janvier 2011).

[4] Voir le site http://www.bluekai.com/exchange.php

[5] « L'exploration des données » peut être définie de nombreuses manières différentes, mais la meilleure définition est peut-être la suivante : large éventail d'activités portant sur des données ; il peut s'agir recherches « subjectives » c'est-à-dire portant sur des individus, ou de recherches « tendancielles » qui mettent en évidence des profils d'activité ou des relations inhabituels ou prédéterminés. Entre ces deux extrémités il y a les recherches « relationnelles » – qui partent d'un individu et vont en s'élargissant pour déterminer qui communique ou interagit avec qui –, et la « mise en correspondance des données », qui consiste à associer deux séries de données ou plus et à rechercher les points de convergence ou de divergence. Cate, F.H. (2008) Government Data Mining: The Need for a Legal Framework, in Harvard Civil Rights-Civil Liberties Law Review, Vol. 43, p. 438.

[6] Compte tenu des différences entre législations nationales en matière de protection de la vie privée, toute extrapolation de ces exemples à d'autres marchés est délicate.

[7] Les données de cartes de crédit proposées sur les marchés de la cybercriminalité sont le numéro de la carte de crédit, la date d'expiration et même, dans certains cas, le nom du titulaire de la carte (ou, le cas échéant, le nom de la société), l'adresse de facturation, le numéro de téléphone, le numéro de sécurité et le code PIN (Symantec, 2010).

[8] On désigne généralement par l'expression « atteinte à la sécurité des données » la perte, la consultation non autorisée ou la divulgation de données de caractère personnel par suite d'un manquement de l'organisation à les protéger efficacement (OCDE, 2011b). Lorsque l'infraction concerne un bien intellectuel autre que des données personnelles, on utilise l'expression « accès non autorisé ».

[9] Les attaques malveillantes demeurent la cause la plus fréquente d'atteintes à la sécurité des données en termes d'enregistrements volés, mais pas en nombre d'incidents. Ainsi, 63 % des enregistrements exposés à des risques suite aux incidents enregistrés entre 2005 et 2011 par *Privacy Rights Clearinghouse* étaient liés à des attaques malveillantes, contre 27 % à des cas de « perte, abandon ou vol d'ordinateur portable, d'assistant personnel, de smartphone, de dispositif de stockage portatif, de CD, de disque dur, de cassette, etc. ». En nombre d'incidents, les attaques malveillantes arrivent en troisième place (19 % de l'ensemble des incidents de ce type), derrière les cas de perte (31 %) et de divulgation non intentionnelle (20 %).

[10] Dans son rapport annuel de 2008, TJX indiquait : « Les coûts réels de l'intrusion informatique pourraient dépasser [les] réserves constituées pour [les] pertes probables estimées, et [notre] réputation et notre activité pourraient être matériellement affectées par une atteinte future à la sécurité des données. ».

[11] Il a été dit que cette attaque pourrait coûter aux seuls émetteurs de cartes bancaires entre 230 et 385 millions USD si tous les clients affectés décidaient de remplacer leur carte. Cette estimation s'appuie sur un coût de remplacement censé être compris entre 3 et 5 USD par carte, et sur un nombre de comptes attaqués de 77 millions (voir Reuters, 2011).

[12] http://mashable.com/2011/01/29/identity-theft-infographic/

[13] AAP (2010) « ID theft costs $1.3bn », *The Australian*, 6 juillet 2010. Disponible à l'adresse : http://www.theaustralian.com.au/australian-it/id-theft-costs-13bn-survey/story-e6frgakx-1225888567247 (septembre 2010).

[14] L'expérimentation économique consiste à étudier le comportement des individus dans des scénarios présentant un intérêt sur le plan économique, au moyen d'incitations pécuniaires. Ces expériences reproduisent certaines conditions des situations réelles dans un cadre expérimental et stylisé.

[15] http://www.experian.com/consumer-products/identity-theft-protection.html

[16] http://www.reputation.com/

[17] http://www.safetyweb.com/