Des **données** très convoitées

Le business des **données personnelles** sur Internet est en plein boom, mais les GAFA (Google, Apple, Facebook, Amazon) trustent 95 % de ce marché. Comment **réguler** le secteur pour protéger la vie privée des citoyens sans tuer la poule aux œufs d'or ?

Tous surveillés. Les traces que l'on laisse derrière soi sur le Net – préférences idéologiques, culinaires ou sportives, achats, peines de cœur, soucis de santé… – sont devenues une industrie extrêmement lucrative. Selon le cabinet IDC, le marché des données des citoyens de l'Europe des Vingt-Huit s'élevait à 60 milliards d'euros en 2016, et devrait atteindre 80 milliards en 2020. Et ce n'est que la partie émergée de l'iceberg. Indirectement, c'est-à-dire en tenant compte du chiffre d'affaires supplémentaire et des emplois générés par les utilisateurs de ces informations, la valeur de *« l'économie européenne des données »* s'élèverait à quelque 300 milliards d'euros en 2016, et pourrait atteindre 430 milliards en 2020. C'est le pétrole du XXIe siècle, assurent les plus enthousiastes. Mais un pétrole qui connaît ses premières marées noires.

Quelques condamnations viennent de rappeler la part d'ombre des données : celle de WhatsApp pour avoir tenté de forcer ses utilisateurs à partager leurs données avec sa maison mère, Facebook. Dans la foulée, celle de Facebook pour avoir juré à la Commission européenne, en 2014, que ce partage de données serait techniquement impossible. Et le 17 mai, c'est la Commission nationale de l'informatique et des libertés (CNIL) qui condamne Facebook pour atteinte à la vie privée, à travers six manquements graves à la loi Informatique et libertés, dont *« la combinaison potentiellement illimitée de toutes les données des utilisateurs… sans qu'ils puissent mettre fin au suivi massif dont ils sont l'objet »*. Les mises en cause des géants du Net ne sont certes pas nouvelles, et les sanctions bien modestes. Mais elles se multiplient, et touchent aujourd'hui au cœur du système : la collecte de ces données personnelles et leur traitement. Les risques que Google, Apple, Facebook ou Amazon font courir à notre société sont jugés tels qu'ils suscitent une volonté de remise à plat musclée des règles par les autorités. Avec pour objectif de redonner aux individus la maîtrise de leurs données personnelles, afin qu'ils sachent enfin ce qui en est fait et qu'ils puissent donner, ou non, leur accord de manière libre et éclairée.

Cette révision met en émoi tous les acteurs qui vivent de cette collecte massive, souvent à l'insu des individus, et qui suivent à la trace leur moindre requête, *like* et autres partages, pour réaliser un profilage commercial de plus en plus fin, dans le but de personnaliser les messages qui leur sont adressés par tous les commerçants et fournisseurs de services pullulant sur la Toile. A commencer par les acteurs de la publicité en ligne, le plus gros consommateur de données personnelles à ce jour. Dans un an, le nouveau règlement général européen sur la protection des données privées, RGDP, entre en vigueur. Il impose une application beaucoup plus stricte des droits existants, avec des sanctions très lourdes, jusqu'à 20 millions d'euros ou 4 % du chiffre d'affaires.

UN MÉTIER QUI SENT LE SOUFRE

« Voilà qui commence à être dissuasif et va contraindre chacun à respecter des règles qui ne l'étaient que rarement », reconnaît Me Alain Bensoussan, spécialiste du droit des technologies. Et le texte apporte quelques changements majeurs : la notion de *« donnée personnelle »* est élargie à *« tout ce qui permet d'identifier directement ou indirectement les personnes »*, y compris l'adresse IP de l'ordinateur et les cookies, ces petits fichiers posés par les collecteurs de données sur le navigateur des internautes pour enregistrer leur activité. Et ces données ne pourront plus être collectées qu'avec le consentement *« libre, éclairé, explicite et univoque »* des individus. En clair, la collecte massive opérée en douce avec un accord donné une fois pour toutes par une case précochée, dite d'*« opt-out »*, et pour des finalités très générales, c'est fini ! Désormais, l'accord devra être acquis par une case *« à cocher »*, dite d'*« opt-in »*, pour des finalités précises et claires, y compris quand les données sont combinées avec d'autres. Il devra être révocable d'un clic, et à tout instant. Et si l'internaute n'y consent pas, l'accès au service ne pourra pas lui être refusé.

L'inquiétude des acteurs est d'autant plus forte que ces obligations s'imposeront à l'avenir à toute la chaîne : du collecteur à l'annonceur, en passant par tous ceux qui manipulent des données personnelles pour *« toucher le bon consommateur au bon moment »*, dans une chaîne ultracomplexe qui comprend aussi des e-mailers, des acheteurs d'espace automatisés, des éditeurs de logiciels de CRM (gestion de la relation client), etc.

Cela concerne bien sûr les collecteurs de données à l'ancienne, éditeurs, sociétés de VPC, sociétés d'annuaires, enseignes émettant des cartes de fidélité, qui tous louent leurs fichiers – comme Orange avec ses 19 millions d'adresses de particuliers, ou la Fnac avec son fichier de 6 millions de clients. Mais aussi les *data brokers*, ces courtiers en données qui agrègent des fichiers et des bases disparates pour vendre des profils à des sociétés. Comme Mediaprism, filiale de La Poste, qui héberge les données de quelque 300 collecteurs et les croise avec ceux de l'Insee, de sites d'e-commerce et de comparateurs Web ; ce qui lui permet d'afficher une base de données portant sur 36 millions de consommateurs dessinés par 150 critères. Mais, surtout, de beaucoup plus gros joueurs, comme les américains eXelate et Acxiom ou le français Weborama, qui créent des audiences cibles en analysant par algorithmes les traces aspirées sur Internet et les réseaux sociaux.

Ces *data brokers* sont déjà dans le viseur de la CNIL, qui a mené 50 contrôles en 2016. D'abord parce que les internautes ne savent souvent rien des données utilisées par les entreprises et de l'usage qui en sera fait. Ils ignorent qu'ils font en réalité l'objet d'un profilage qui n'est plus seulement sociologique, mais de plus en plus comportemental – distinguer par exemple les bons et les mauvais payeurs –, et qui peut être à l'origine d'une discrimination invisible. Ensuite parce que la CNIL a pu constater *« la difficulté à apporter la preuve d'un consentement éclairé, des finalités initiales de collecte détournées lors de la combinaison avec d'autres bases de données, ou des durées de conservation des données beaucoup trop longues »*. Bref, le métier sent le soufre.

LE NOUVEAU REGLEMENT EUROPÉEN SUR LA PROTECTION DE LA VIE PRIVÉE IMPOSERA UNE APPLICATION BEAUCOUP PLUS STRICTE DES DROITS EXISTANTS

La principale source d'inquiétude des régulateurs concerne les GAFA (Google, Apple, Facebook, Amazon), qui dominent de la tête et des épaules le monde du numérique.

L'effet de réseau, par lequel plus un service est utilisé plus il a de valeur pour l'utilisateur, a fini par tuer toute concurrence. Les GAFA verrouillent la collecte, les algorithmes de ciblage comportemental en temps réel et l'espace publicitaire. Ils captent ainsi 68 % du marché de la pub en ligne en France (77 % au Royaume-Uni), et toute sa croissance, selon l'Interactive Advertising Bureau (IAB), qui fédère les acteurs de la publicité sur Internet. Cette intégration verticale totale et fermée aux tiers leur donne une puissance absolue. *« Le marché du data est dominé à 95 % par les GAFA, tous les autres acteurs réunis n'en détiennent pas plus de 5 % »,* confirme Alain Lévy, PDG de Weborama, membre de l'IAB France. *« Google peut voir ce que les gens cherchent, Facebook ce qu'ils aiment, Amazon ce qu'ils achètent »,* écrivait en mai *The Economist,* pointant cet *« œil divin »* sur toute l'économie mondiale. *« Qu'un produit marche, et ils piocheront dans leurs ressources quasi illimitées pour le racheter. »* L'autorité française de la concurrence doit proposer cet été une réforme des règles de la publicité en ligne. Mais cela suffira-t-il à ouvrir le secteur ?

Quant à la taille de ces bases de données, est-elle encore compatible avec la protection de la vie privée ? Avec l'explosion des traces laissées sur Internet et les réseaux sociaux, et la combinaison des données pratiquée par toutes les plates-formes, les mégabases dominent le marché. *« Le croisement massif de données »* a un *« caractère particulièrement intrusif »,* souligne la CNIL dans sa condamnation de Facebook. Et elle poursuit : *« La combinaison potentiellement illimitée de toutes les données des utilisateurs collectées non seulement sur Facebook mais sur des sites et applications tiers (...) est de nature à porter atteinte à leur vie privée. »*

Même des *data brokers* modestes sont contraints de se lancer dans la course au gigantisme des bases pour affiner leur profilage. La trajectoire du français Weborama est symptomatique de la transformation du métier. Créée en 1998, l'entreprise était à l'origine spécialisée dans la mesure d'audience des sites, puis s'est développée dans l'achat d'espaces. Rachetée par Alain Lévy en 2005, elle a opéré un recentrage radical sur l'analyse des données : *« Avant, on vendait le profil avec le média,* commente le PDG. *Mais le marché a évolué, et la valeur ajoutée s'est concentrée dans le traitement des données. Désormais, on vend des profils sans le média pour aider nos clients à faire évoluer leur stratégie autour de la compréhension scientifique du consommateur. »* Weborama crée ainsi des profils à partir de l'analyse sémantique des mots utilisés sur les réseaux sociaux. Il a répertorié 200 segments sur pas moins de 850 millions d'internautes (dont 300 millions aux Etats-Unis). Son métier : vendre des profils construits par algorithmes. Ce positionnement, qui lui a permis de retrouver de la croissance mais qui est très coûteux en investissements, est le reflet d'un métier contraint de se déplacer vers toujours plus de technologie et de plonger dans le big data.

Et ce n'est pas fini. Car les acteurs les plus en pointe sont en passe de franchir une nouvelle frontière : suivre les consommateurs non plus seulement sur Internet, mais aussi dans leurs déplacements physiques en magasins ou quand ils interrogent des opérateurs de centres d'appels. Et ce, en reliant toutes leurs données à un identifiant unique, comme l'explique Julien Hirth dans *Le Data marketing* (Eyrolles, 294 p., 25 €). Dans le métier, on appelle cela du *« CRM onboarding »,* un service vendu par l'américain Acxiom, qui revendique une base de plusieurs dizaines de millions d'internautes français. Son premier client en France ? Carrefour, qui cherchait à compléter la connaissance de ses clients pour mieux cibler ses offres promotionnelles et permettre aux marques de s'adresser à eux par son intermédiaire.

Jusqu'où ces *data companies* iront-elles dans la course au gigantisme et au perfectionnement du traçage ? Parmi les internautes, beaucoup ne croient que modérément aux promesses de respect de l'anonymat. *« Quand on dispose de tant d'informations sur les individus, retirer les identifiants ne suffit plus à garantir leur anonymat,* considère Yves-Alexandre de Montjoye, spécialiste de la question au Data Science Institute de l'Imperial College de Londres. *Il suffit de quelques points pour les réidentifier. Et cela, les personnes concernées ne le savent pas. Seules des données agrégées permettent de garantir l'anonymat. »*

UNE DIMENSION GÉOPOLITIQUE

L'Europe ayant perdu la bataille des bases de données, c'est dans les algorithmes que de nombreuses start-up cherchent leur voie, en particulier en France, pays de mathématiciens et de *data scientists.* Ainsi de Zettafox, créée en 2008, qui construit des profils à partir de l'historique des données personnelles de ses clients. Point de mégabases, juste du calcul. La start-up vend, comme d'autres acteurs, son modèle d'*« analyse prédictive »* qui *« score les profils ».* Mais Zettafox a aussi développé une *« analyse prescriptive »,* bâtie sur des scénarios, visant à comprendre la logique interne du comportement des clients. *« Notre objectif est d'arriver à fournir une analyse en temps réel du comportement des individus pour permettre à nos clients de changer de stratégie aussi vite que le marché »,* déclare Marc Attalah chez Zettafox.

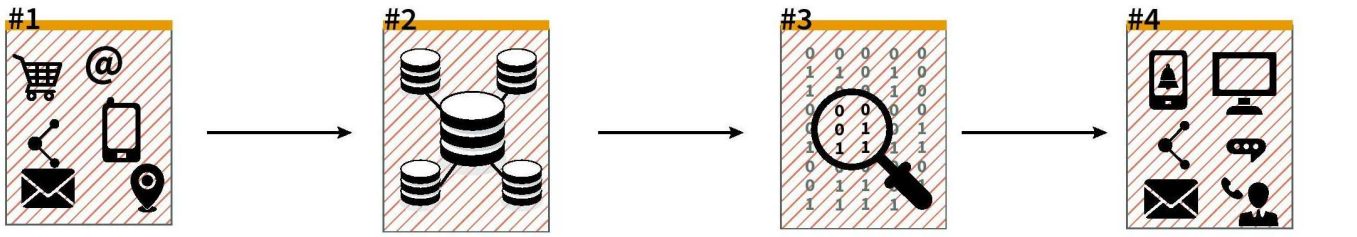
Une stratégie qui n'efface pas toutes les inquiétudes. Car la course à la collecte est si stratégique pour bâtir des algorithmes et nourrir l'intelligence artificielle que l'affaire a pris une dimension géopolitique : *« Si on nous bloque aujourd'hui en Europe avec un RGDP trop strict,* dit Marc Attalah, *on ne pourra plus accéder à la masse de données nécessaire pour bâtir cette nouvelle génération de services bâtis sur l'intelligence artificielle et espérer concurrencer un jour Google et Facebook. Ils auront alors construit une avance irrattrapable, sauf peut-être par WeChat et Baidu, car la Chine est moins regardante sur les questions de vie privée. »*

Aussi la gestion du consentement des utilisateurs est-elle devenue de la plus haute importance. Dans ce domaine aussi, des start-up se développent, comme la française Crystalchain qui, selon un de ses fondateurs, Pierre Achach, *« doit permettre au consommateur de reprendre le contrôle de ses données sur un tableau de bord hébergé sur une plate-forme en marque blanche de gestion des consentements reposant sur la technologie sécurisée de la* blockchain ». Le salut des utilisateurs viendra autant de la technologie que de la réglementation. ■

PARMI LES INTERNAUTES, BEAUCOUP NE CROIENT QUE MODEREMENT AUX PROMESSES DE RESPECT DE L'ANONYMAT

Les données, matière première d'un business florissant

Le parcours des données



COLLECTE ET VENTE

Données privées récoltées principalement sur Internet via des connexions sur les sites (e-commerce et autres), des applications, des publications sur les réseaux sociaux et des campagnes d'e-mailing, etc.

STOCKAGE ET STRUCTURATION

Intégration, hébergement et gestion de bases de données.

ANALYSE ET INTELLIGENCE

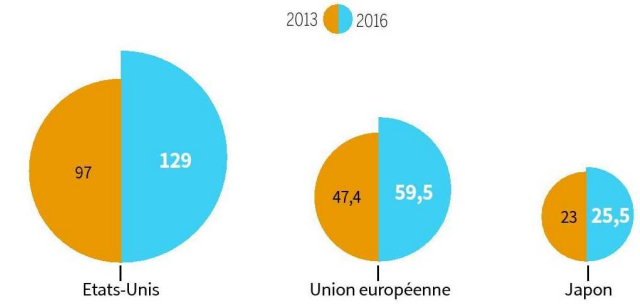
Exploration (datamining) et traitement des données (data science) afin d'en extraire des informations exploitables.

DIFFUSION ET CIBLAGE

Utilisation des informations compilées pour augmenter l'efficacité du ciblage publicitaire, des logiciels d'intelligence artificielle et autres services commerciaux.

Un marché en expansion

CHIFFRE D'AFFAIRES DES ENTREPRISES SPÉCIALISÉES DANS LES SOLUTIONS BIG DATA, EN MILLIARDS D'EUROS

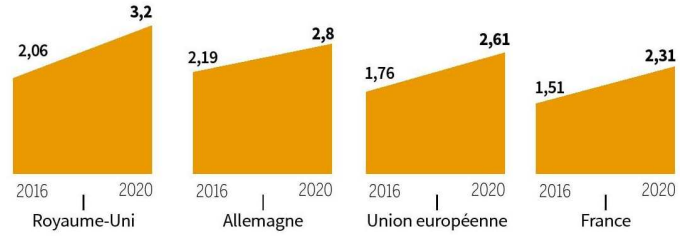


NOMBRE D'ENTREPRISES PRÉSENTES SUR LE MARCHÉ, EN MILLIERS

Entreprises spécialisées dans les solutions big data

Entreprises utilisant des informations obtenues via des solutions big data

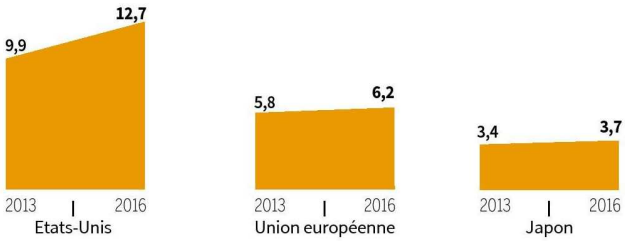
POIDS DE L'ÉCONOMIE DES DONNÉES*, EN % DU PIB



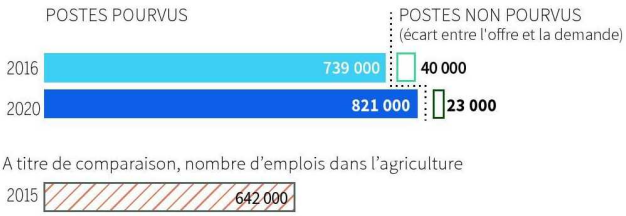
* Impacts économiques directs, indirects et induits

Des compétences recherchées

NOMBRE D'EMPLOIS SPÉCIALISÉS EN SOLUTIONS BIG DATA, EN MILLIONS



NOMBRE D'EMPLOIS SPÉCIALISÉS EN SOLUTIONS BIG DATA, EN FRANCE



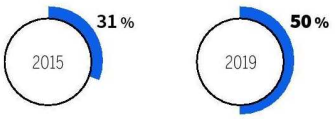
A titre de comparaison, nombre d'emplois dans l'agriculture

Le big data au service de la publicité

INVESTISSEMENTS EN PUBLICITÉ PROGRAMMATIQUE*, EN MILLIONS D'EUROS, EN 2015

* Adaptation de la publicité au public grâce à des algorithmes et l'analyse de big data

PART DE LA PROGRAMMATIQUE DANS L'ENSEMBLE DES INVESTISSEMENTS PUBLICITAIRES MONDIAUX



Note : Les solutions big data englobent la production et l'offre de produits, services et technologies permettant la collecte, la structuration, le stockage, l'exploitation et l'analyse de grands ensembles de données brutes pour en extraire des informations

SOURCES « EUROPEAN DATA MARKET », IDC ET OPEN EVIDENCE POUR LA COMMISSION EUROPÉENNE, MAI 2017 ; « DATA DRIVEN INNOVATION. BIG DATA FOR GROWTH AND WELL BEING », OCDE, 2015 ; *LE MARKETING DES DATAS*, JULIEN HIRTH, EYROLLES, 2017 ; LIMELIGHT CONSULTING ET MEDIAPOST COMMUNICATION ; MAGNA GLOBAL - **INFOGRAPHIE** *ÉLISA BELLANGER, MARIANNE BOYER, LÉA DESRAYAUD, VALÉRIE SEGOND*

Isabelle Falque-Pierrotin : « La fin du chèque en blanc »

PRÉSIDENTE DE LA CNIL et du G29, qui réunit tous les régulateurs européens, Isabelle Falque-Pierrotin explique les enjeux du cadre européen de protection des données qui entrera en vigueur en mai 2018.

Que signifie ce nouveau règlement pour les particuliers ?

C'est la fin du chèque en blanc sur les données. Les consommateurs ne veulent plus que leurs données soient pillées. Aujourd'hui, la dissymétrie est trop grande entre les GAFA *[Google, Apple, Facebook, Amazon]* et les citoyens. Ce règlement permet de renforcer les droits des individus par rapport aux grands acteurs économiques de l'Internet. L'esprit, c'est de créer un marché européen des données, de donner une vision commune de la protection des données personnelles. Ce règlement s'inscrit dans une double inspiration : alléger considérablement le contrôle a priori et responsabiliser les responsables de traitement, qui doivent prouver qu'ils protègent correctement les données.

Mais le ciblage, qui sera plus encadré, permet de rendre la publicité moins intrusive et donc plus utile pour les utilisateurs…

Il y aura un équilibre à trouver entre les besoins des acteurs et le désir des consommateurs de reprendre la main sur leurs données. Ce n'est pas pour rien que l'on a vu les bloqueurs de publicité se développer fortement. Demain, nous aurons la capacité d'aller voir ces moissonneuses de données et d'ouvrir leur capot pour comprendre comment cela fonctionne.

Cette réglementation ne risque-t-elle pas de nuire à la compétitivité des entreprises européennes ?

Je ne le crois pas. Depuis l'affaire Snowden, il y a un déficit de confiance vis-à-vis de certains services numériques. Nous sommes tous consommateurs de ces services, mais les individus demandent de plus en plus à maîtriser leurs données, toutes les enquêtes le montrent. L'Europe a fait le choix de s'inscrire dans cette tendance et d'anticiper un mouvement qui sera mondial. Elle se donne les moyens d'une innovation plus durable. Depuis deux ou trois ans, on assiste au développement de *dashboards [tableaux de bord],* qui permettent de mieux paramétrer et contrôler ses données. Certaines start-up intègrent la protection de la vie privée dès la conception de leur produit ou en font un argument de vente.

Les géants du Net n'en sortiront-ils pas favorisés par rapport aux entreprises européennes ?

Au contraire. Ce règlement remet les acteurs internationaux à égalité : même s'ils ne sont pas établis en Europe, dès lors qu'ils ciblent un utilisateur européen, la loi européenne s'applique. D'autant que le règlement prévoit des sanctions allant jusqu'à 4 % du chiffre d'affaires. C'est bien plus élevé que ce qui existe actuellement. Aujourd'hui, les amendes sont ridicules et nous avons des discussions sans fin avec Facebook ou Google pour savoir quelle réglementation – nationale ou américaine – doit s'appliquer. C'est un énorme sujet de contentieux.

Les entreprises sont-elles prêtes ?

Les entreprises ne sont pas préparées. Pour l'instant, elles ont simplement identifié qu'il s'agissait d'un coût. Elles ne se rendent pas compte que ce coût est aussi un investissement. C'est un changement de culture : elles doivent considérer que la protection des données n'est plus une contrainte juridique, mais un élément consubstantiel à leur activité. C'est une dimension à intégrer au niveau du marketing, des ressources humaines, de la recherche-développement, de la relation client. Elles doivent repenser leur gouvernance, savoir si elles ont les compétences en interne, le bon système d'information… Et, enfin, articuler la protection des données avec la cybersécurité. Les enjeux de protection des données doivent remonter au niveau des décisions stratégiques, être arbitrés au niveau des comités de direction.

Cela va coûter cher pour les PME-PMI…

Elles sont aujourd'hui très loin de ces problématiques. Nous réfléchissons à un outil clés en main, un pack de conformité pour les aider.

Ce règlement signe-t-il la mort des autorités de régulation locales ?

Pas du tout. Il vise plutôt à renforcer les autorités nationales, qui vont pouvoir coopérer et mettre en place des investigations communes sur certains sujets. Ce type de fonctionnement sera très nouveau en Europe. ■

PROPOS RECUEILLIS PAR SANDRINE CASSINI ET MARTIN UNTERSINGER