Collecte de données et profilage

Comment les entreprises savent tout de vous

Stéphane FOSSE
fosse.fr

04 avril 2026

Résumé

En 2017, le data broker Acxiom détenait plus de 5 000 attributs sur environ 700 millions de consommateurs à travers le monde. Oracle, de son côté, revendiquait plus de 30 000 segments comportementaux construits à partir de deux milliards de profils. Ces chiffres, tirés d'une enquête indépendante du chercheur Wolfie Christl [3], illustrent une réalité que la plupart des citoyens ignorent : une industrie entière vit de la collecte, du croisement et de la revente de leurs données personnelles, sans qu'ils en aient jamais entendu parler.

Qu'est-ce qu'un data broker et combien de données possède-t-il sur chaque consommateur ?

Un data broker est une entreprise dont l'activité principale consiste à collecter des informations personnelles sur les consommateurs auprès de sources multiples, à les agréger, les analyser et les revendre à des tiers. La Federal Trade Commission (FTC) américaine a formalisé cette définition en 2014 dans son rapport *Data Brokers : A Call for Transparency and Accountability* [6], après avoir enquêté sur neuf acteurs représentatifs du secteur : Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf et Recorded Future. Ces neuf entreprises généraient à elles seules 426 millions de dollars de revenus annuels en 2012.

Le trait distinctif de ces entreprises, c'est qu'elles n'interagissent jamais directement avec les personnes dont elles collectent les données. Un consommateur peut vivre toute sa vie sans jamais entendre le nom d'Acxiom ou d'Epsilon, alors que ces sociétés détiennent sur lui des informations sur son affiliation religieuse, ses opinions politiques, ses habitudes d'achat, son état de santé supposé, sa situation financière, la composition de son foyer, son type de véhicule, et même la probabilité qu'il change d'emploi dans les mois à venir.

La commission Commerce du Sénat américain a documenté en décembre 2013 qu'un seul data broker pouvait détenir jusqu'à 75 000 éléments de données distincts dans ses systèmes [11]. Le rapport parlementaire décrivait des catégories de segmentation aux intitulés révélateurs : « Rural and Barely Making It », « Ethnic Second-City Strugglers », « Retiring on Empty : Singles » ou encore « Tough Start : Young Single Parents ». Ces étiquettes ne sont pas des curiosités sociologiques. Elles servent à cibler des populations financièrement vulnérables pour leur vendre des prêts à taux élevé, des assurances coûteuses ou des produits financiers à risque.

Comment les entreprises collectent-elles nos données personnelles au quotidien ?

Les sources de données exploitées par les data brokers se répartissent en trois grandes familles, selon la classification de la FTC : les sources gouvernementales, les sources publiquement accessibles et les sources commerciales. Aucune de ces sources, prise isolément, ne semble menaçante. C'est leur agrégation qui produit un portrait d'une précision inquiétante.

Les registres publics et gouvernementaux. Les États alimentent eux-mêmes cette industrie, souvent sans en mesurer les conséquences. Au Royaume-Uni, Acxiom liste parmi ses sources le cadastre (HM Land Registry), l'Office for National Statistics (qui conduit les recensements), le ministère de l'Industrie et l'autorité de régulation des télécommunications Ofcom, comme l'a documenté Privacy International en 2018 [8]. Aux États-Unis, la FTC a constaté que les data brokers puisent dans les fichiers du recensement fédéral, les registres de propriété, les listes électorales (avec l'affiliation partisane), les fichiers de permis de conduire, les casiers judiciaires, les registres de mariages et de divorces, les licences professionnelles et même les permis de chasse et de pêche. En France, le registre électoral est en principe protégé, mais de nombreuses bases publiques restent accessibles.

Le pistage sur le web : cookies, pixels et empreintes numériques. La plupart des sites web intègrent du code tiers invisible qui collecte des informations sur les visiteurs. Les cookies tiers, déposés par des réseaux publicitaires, permettent de suivre un internaute de site en site et de reconstituer son historique de navigation. L'agence européenne ENISA a documenté dès 2012 que les dix plus grands trackers tiers étaient présents sur plus de 70 % des sites web populaires, un chiffre en hausse constante depuis 2005 [2]. Le pistage par e-mail est tout aussi répandu : une étude citée par Privacy International estimait en 2016 que plus de 40 % des e-mails envoyés dans le monde contenaient un pixel de suivi, une image invisible de 1×1 pixel qui signale à l'expéditeur quand le message est ouvert, depuis quel appareil et dans quel lieu géographique.

Au-delà des cookies, le *browser fingerprinting* (empreinte de navigateur) permet d'identifier un internaute sans déposer le moindre fichier sur son appareil. L'ENISA a relevé qu'en combinant la version du navigateur, les polices installées, la résolution d'écran et d'autres paramètres techniques, il était possible d'identifier un navigateur parmi 290 000 autres avec une forte probabilité. Ce type de pistage, dit « sans état », est redoutable parce qu'il est quasi indétectable par l'utilisateur.

Les applications mobiles et leurs SDK. Plus de trois applications Android sur quatre contiennent au moins un tracker tiers, selon les recherches rapportées par Privacy International [8]. Le mécanisme repose souvent sur des SDK (*Software Development Kits*) fournis gratuitement aux développeurs par les data brokers. En échange de fonctionnalités techniques, le SDK aspire des données de géolocalisation, de contacts, de photos, voire d'usage du micro. La FTC a documenté en 2024, dans son action contre le data broker X-Mode (rebaptisé Outlogic), que ce type de collecte permettait d'amasser dix milliards de points de géolocalisation par jour, avec une précision de vingt mètres, et que ces données étaient revendues à des contractants du gouvernement américain.

Les achats en magasin, les cartes de paiement et les programmes de fidélité. Quand un consommateur utilise sa carte bancaire, les données agrégées de sa transaction alimentent les profils des data brokers. Visa et MasterCard figurent toutes deux comme fournisseurs de données dans le répertoire d'Oracle, selon l'enquête de Wolfie Christl publiée en 2017 [3]. MasterCard se distingue en associant ses données transactionnelles à des identifiants de cookies, ce qui permet de relier les achats physiques au profil numérique d'un individu. La société revendiquait des données issues de 2,2 milliards de cartes de paiement et 43 milliards de transactions annuelles.

Les enquêtes, jeux-concours et quiz en ligne. Chaque formulaire rempli, chaque participation à un tirage au sort est une occasion de capturer des données. Le data broker Epsilon a construit une base de données appelée *Shopper's Voice* qui couvre 1 000 attributs par consommateur, récoltés via des enquêtes en échange de bons de réduction et de la possibilité de gagner 1 500 dollars. C'est le même ressort qui a rendu possible le scandale Cambridge Analytica en 2018 : des quiz de personnalité conçus pour extraire des données à grande échelle.

Le pistage physique. La collecte ne s'arrête pas au monde numérique. Des entreprises comme Freckle IoT déploient des balises Bluetooth et des capteurs dans les magasins, restaurants, aéroports, centres commerciaux et même le mobilier urbain pour suivre les déplacements des consommateurs en temps réel via leur téléphone. Freckle, intégré dans plus de 2 000 applications mobiles, fournissait ces données de localisation à Acxiom via sa plateforme LiveRamp.

Comment ces données sont-elles croisées et transformées en profils exploitables ?

La puissance des data brokers ne réside pas dans la collecte brute mais dans le croisement. La FTC a détaillé en 2014 un processus en trois étapes que l'industrie appelle *onboarding* : segmentation, appariement et ciblage [6]. D'abord, les données offline (achats en magasin, registres publics, enquêtes) sont associées à des identifiants numériques via des cookies ou des identifiants d'appareil mobile. Ensuite, ces profils enrichis sont comparés aux bases de données des sites web où le consommateur est inscrit. Enfin, le profil complet est mis à disposition des annonceurs pour du ciblage en temps réel.

Le système *Identity Graph* d'Oracle illustre bien cette mécanique. Il promet de « unifier les identités adressables sur tous les appareils, écrans et canaux » en reliant adresses postales, adresses e-mail, comptes utilisateur, identifiants mobiles, identifiants de box TV et identifiants de cookies en un seul profil. Acxiom, de son côté, a développé le système Abilitec qui fait converger noms, adresses, e-mails, numéros de téléphone, adresses IP et données de géolocalisation vers un identifiant persistant unique par individu. Grâce à sa plateforme LiveRamp, acquise en 2014, Acxiom connecte ensuite ces profils à plus de 500 plateformes marketing et à plus de 100 fournisseurs de données tiers, dont Equifax, Experian et TransUnion — les trois principales agences de crédit américaines.

Le résultat, c'est un système où sept des neuf data brokers étudiés par la FTC achètent et vendent des données les uns aux autres, rendant quasiment impossible pour un consommateur de remonter à la source d'une information le concernant. L'enquête parlementaire du Sénat américain comparait ce circuit à des poupées russes : ouvrir un data broker révèle des centaines de data brokers plus petits à l'intérieur.

Pourquoi l'argument « je n'ai rien à cacher » ne tient pas face au profilage massif ?

L'argument est aussi répandu qu'il est mal calibré. Le profilage commercial ne cherche pas des secrets. Il cherche des *patterns* — des régularités comportementales qui permettent de prédire ce que vous ferez demain et d'en tirer profit aujourd'hui.

Shoshana Zuboff, professeure émérite à la Harvard Business School et autrice du livre *The Age of Surveillance Capitalism* publié en 2019 [12], a forgé le concept de « surplus comportemental » (*behavioral surplus*) pour décrire ce phénomène. Les données collectées dépassent ce qui serait nécessaire pour améliorer un service. Ce surplus est transformé en « produits de prédiction » vendus sur ce que Zuboff appelle des « marchés de comportements futurs ». Les acheteurs ? Des annonceurs, des assureurs, des banques, des employeurs — toute entreprise ayant un intérêt commercial à savoir ce que vous allez faire. Dans une interview accordée à la Harvard Gazette en 2019 [7], Zuboff précisait que la dynamique concurrentielle du capitalisme de surveillance a poussé les entreprises au-delà de la simple observation vers ce que les data scientists appellent l'« actuation » : la modification active des comportements par des incitations subtiles, des récompenses et des sanctions invisibles.

Concrètement, cela signifie que votre historique d'achat de paracétamol, combiné à votre code postal et à votre statut matrimonial, peut vous placer dans la catégorie « Allergy Sufferer » ou « Cholesterol Focus » dans les bases d'un data broker — des catégories que la FTC a effectivement identifiées en 2014. Ces étiquettes déterminent les publicités que vous voyez, le prix qui vous est affiché sur un site de voyage, et potentiellement les conditions d'un prêt bancaire. Le data broker InfoUSA a vendu des listes intitulées « Suffering Seniors » à des individus qui les ont utilisées pour cibler des personnes âgées avec des arnaques, selon le New York Times.

Le plus troublant dans ce mécanisme, c'est que des données apparemment insignifiantes deviennent significatives par croisement. Privacy International a formulé ce principe avec clarté [9] : le résultat de l'analyse dépasse la somme de ses parties. Des états émotionnels comme la confiance, la nervosité ou la fatigue peuvent être inférés à partir de simples schémas de frappe au clavier. Des données que personne ne qualifierait de « sensibles » prises isolément — l'heure à laquelle vous ouvrez une application, la fréquence de vos déplacements dans un quartier donné, le type de musique que vous écoutez — deviennent, une fois croisées, un portrait intime de votre situation personnelle, financière, médicale et politique.

Dean Curran, chercheur en sociologie à l'Université de Calgary, a poussé l'analyse plus loin dans un article publié en 2023 dans *Big Data & Society* [4]. Selon Curran, le capitalisme de surveillance génère un double impératif : collecter toujours plus de données (*imperative to collect*) et connecter toujours plus de systèmes entre eux (*imperative to connect*). Ce double impératif ne produit pas seulement une surveillance invasive. Il crée aussi une fragilité systémique : quand ces réseaux massivement interconnectés tombent en panne ou sont compromis, les conséquences se propagent en cascade à travers toute la vie sociale et économique. Les fuites de données massives chez Facebook et Twitter ne sont pas des accidents isolés — elles sont des conséquences structurelles d'un modèle fondé sur l'accumulation maximale.

Quel cadre juridique protège les citoyens européens contre le profilage commercial ?

Le Règlement général sur la protection des données (RGPD), en vigueur dans l'Union européenne depuis le 25 mai 2018, a introduit une définition explicite du profilage à l'article 4(4) : « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique. » Le Groupe de travail Article 29 (devenu depuis le Comité européen de la protection des données) a précisé en octobre 2017 les lignes directrices applicables [1], insistant sur les principes de transparence, de minimisation et de base légale.

En théorie, le RGPD donne aux citoyens européens des droits substantiels : droit d'accès, droit de rectification, droit d'opposition au profilage. En pratique, le fossé reste considérable. En novembre 2018, Privacy International a déposé des plaintes formelles contre Acxiom et Oracle auprès de l'Information Commissioner's Office (ICO) britannique [9]. L'organisation a documenté que ces entreprises profitent de l'opacité de leurs opérations pour contourner les obligations de transparence, que les réponses aux demandes d'accès des personnes concernées étaient lacunaires, et que les bases légales invoquées (consentement ou intérêt légitime) ne résistaient pas à l'examen.

Aux États-Unis, la situation est encore plus ouverte. Le rapport du GAO commandé par le Sénat en 2013 concluait qu'aucune loi fédérale ne donnait aux consommateurs le droit de savoir quelles informations les data brokers détenaient sur eux à des fins marketing, ni le droit de corriger ces informations. Le *Fair Credit Reporting Act* ne couvre que les données utilisées pour des décisions de crédit, d'emploi ou d'assurance — pas le marketing, qui représente pourtant la majeure partie de l'activité des data brokers.

L'ENISA, l'agence européenne de cybersécurité, plaidait dès 2015 dans son rapport *Privacy by design in big data* [5] pour que la protection de la vie privée soit intégrée dès la conception des architectures de

données — une approche de *privacy by design* qui devrait être la norme pour tout système traitant des données personnelles à grande échelle. Cette recommandation reste, huit ans plus tard, davantage un vœu qu'une réalité industrielle.

Ce que vous pouvez faire

La première étape est la prise de conscience — et c'est le but de cet article. Tant que les data brokers restent des noms inconnus du grand public, ils prospèrent dans l'ombre. La deuxième étape est d'exercer ses droits : en Europe, le RGPD donne à chaque citoyen le droit de demander à n'importe quelle entreprise quelles données elle détient sur lui (article 15) et de s'opposer au profilage (article 21). En pratique, cela signifie écrire aux data brokers identifiés dans cet article et exiger l'accès puis la suppression de vos données. Les réponses obtenues — ou leur absence — sont en elles-mêmes instructives.

Sur le plan technique, quelques mesures limitent l'exposition : refuser systématiquement les cookies non essentiels, utiliser un bloqueur de publicités et de trackers, vérifier les permissions accordées aux applications mobiles, préférer des moteurs de recherche et des navigateurs qui ne profilent pas. Ces mesures ne suppriment pas le problème — elles en réduisent l'ampleur.

Mais la réponse de fond est politique et collective. Comme le soulignait Shoshana Zuboff, trois chantiers sont nécessaires pour sortir de cette ère : un changement d'opinion publique qui refuse la normalisation de ces pratiques, une réponse législative qui dépasse le cadre actuel du RGPD et du *Fair Credit Reporting Act* pour réglementer spécifiquement les data brokers, et l'émergence d'alternatives technologiques et commerciales fondées sur le respect de la vie privée. Urbano Reviglio, dans un article publié en 2022 dans *Internet Policy Review* [10], plaidait pour un registre obligatoire des data brokers, des définitions juridiques harmonisées à l'échelle internationale et la création d'une autorité de supervision dédiée — car le RGPD seul ne suffit pas à démanteler une industrie qui génère plus de 200 milliards de dollars de revenus annuels.

Conclusion

L'industrie des data brokers prospère sur un déséquilibre fondamental : ces entreprises savent tout de nous, et nous ne savons rien d'elles. Elles connaissent nos habitudes d'achat, nos fragilités financières, nos centres d'intérêt, nos déplacements, nos relations, et parfois nos états émotionnels — sans jamais nous avoir rencontrés, sans jamais nous avoir demandé quoi que ce soit. Le rapport de la FTC de 2014, les enquêtes du Sénat américain, les plaintes de Privacy International, les recherches de Wolfie Christl et de Shoshana Zuboff convergent vers le même constat : la collecte de données personnelles n'est pas un effet secondaire de l'économie numérique. C'est son moteur.

À ceux qui se disent qu'ils n'ont rien à cacher, la question à poser est celle-ci : êtes-vous d'accord pour que des entreprises dont vous n'avez jamais entendu le nom décident quels produits vous seront proposés, à quel prix, et avec quel niveau de risque — sur la base d'un profil que vous n'avez jamais vu, que vous ne pouvez pas corriger, et dont vous ne pouvez même pas vérifier l'exactitude ?

Références

[1] Article 29 Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Anglais. WP251rev.01. European Commission, oct. 2017.

[2] Claude Castelluccia et Arvind Narayanan. Privacy considerations of online behavioural tracking. Anglais. ENISA, 2012.

[3] Wolfie Christl. Corporate Surveillance in Everyday Life: How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions. Anglais. Cracked Labs, juin 2017.

[4] Dean Curran. Surveillance capitalism and systemic digital risk: The imperative to collect and connect and the risks of interconnectedness. Anglais. In : *Big Data & Society* 10.1 (2023).

[5] Giuseppe D'Acquisto et al. Privacy by design in big data. Anglais. ENISA, déc. 2015.

[6] Federal Trade Commission. Data Brokers: A Call for Transparency and Accountability. Anglais. Federal Trade Commission, mai 2014.

[7] John Laidler. High tech is watching you. Anglais. Mars 2019.

[8] Privacy International. How do data companies get our data? Anglais. Mai 2018.

[9] Privacy International. Submission to the Information Commissioner: Request for an Assessment Notice of Data Brokers Acxiom & Oracle. Anglais. Privacy International, nov. 2018.

[10] Urbano Reviglio. The untamed and discreet role of data brokers in surveillance capitalism: a transnational and interdisciplinary overview. Anglais. In : *Internet Policy Review* 11.3 (2022).

[11] US Senate Committee on Commerce, Science, and Transportation. A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes. Anglais. US Senate, déc. 2013.

[12] Shoshana Zuboff. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Anglais. PublicAffairs, 2019. isbn : 978-1-61039-569-4.