

# Ce que récolte Google : revue de détail

Le temps n'est plus où il était nécessaire d'alerter sur la prédation opérée par Google et ses nombreux services sur nos données personnelles. Il est fréquent aujourd'hui d'entendre dire sur un ton fataliste : « de toute façon, ils espionnent tout »

Si beaucoup encore proclament à l'occasion « je n'ai rien à cacher » c'est moins par conviction réelle que parce que chacun en a fait l'expérience : « on ne peut rien cacher » dans le monde numérique. Depuis quelques années, les mises en garde, listes de précautions à prendre et solutions alternatives ont été largement exposées, et Framasoft parmi d'autres y a contribué.

Il manquait toutefois un travail de fond pour explorer et comprendre, une véritable étude menée suivant la démarche universitaire et qui, au-delà du jugement global approximatif, établisse les faits avec précision.

C'est à quoi s'est attelée l'équipe du professeur Douglas C. Schmidt, spécialiste depuis longtemps des systèmes logiciels, chercheur et enseignant à l'Université Vanderbilt, qui livre au public une étude d'une cinquantaine de pages intitulée Google Data Collection. Cette étude, qui nous semble pouvoir servir de référence, a retenu l'attention du groupe Framalang qui vous en livre ci-dessous l'*executive summary*, c'est-à-dire une sorte de résumé initial, qui en donne un aperçu programmatique.

Si vous trouvez un intérêt à cette traduction et souhaitez que Framalang vous propose la suite nous ferons de notre mieux...

**Traduction Framalang : Alain, fab, FranBAG, Goofy, jums, Khrys, Mika, Piup, serici**

# La collecte de données de Google

## Un premier aperçu

1.

Google est la plus grosse agence de publicité numérique du monde <sup>1</sup>. Elle fournit aussi le leader des navigateurs web <sup>2</sup>, la première plateforme mobile <sup>3</sup> ainsi que le moteur de recherche le plus utilisé au monde <sup>4</sup>. La plateforme vidéo de Google, ses services de mail et de cartographie comptent 1 milliard d'utilisateurs mensuels actifs chacun <sup>5</sup>. Google utilise l'immense popularité de ses produits pour collecter des données détaillées sur le comportement des utilisateurs en ligne comme dans la vie réelle, données qu'il utilisera ensuite pour cibler ses utilisateurs avec de la publicité payante. Les revenus de Google augmentent significativement en fonction de la finesse des technologies de ciblage des données.

2.

Google collecte les données utilisateurs de diverses manières. Les plus évidentes sont « actives », celles dans lesquelles l'utilisateur donne directement et consciemment des informations à Google, par exemple en s'inscrivant à des applications très populaires telles que YouTube, Gmail, ou le moteur de recherche. Les voies dites « passives » utilisées par Google pour collecter des données sont plus discrètes, quand une application devient pendant son utilisation l'instrument de la collecte des données, sans que l'utilisateur en soit conscient. On trouve ces méthodes de collecte dans les plateformes (Android, Chrome), les applications (le moteur de recherche, YouTube, Maps), des outils de publication (Google Analytics, AdSense) et de publicité (AdMob, AdWords). L'étendue et l'ampleur de la collecte passive de données de Google ont été en grande partie négligées par les études antérieures sur le sujet <sup>6</sup>.

3.

Pour comprendre les données que Google collecte, cette étude s'appuie sur quatre sources clefs :

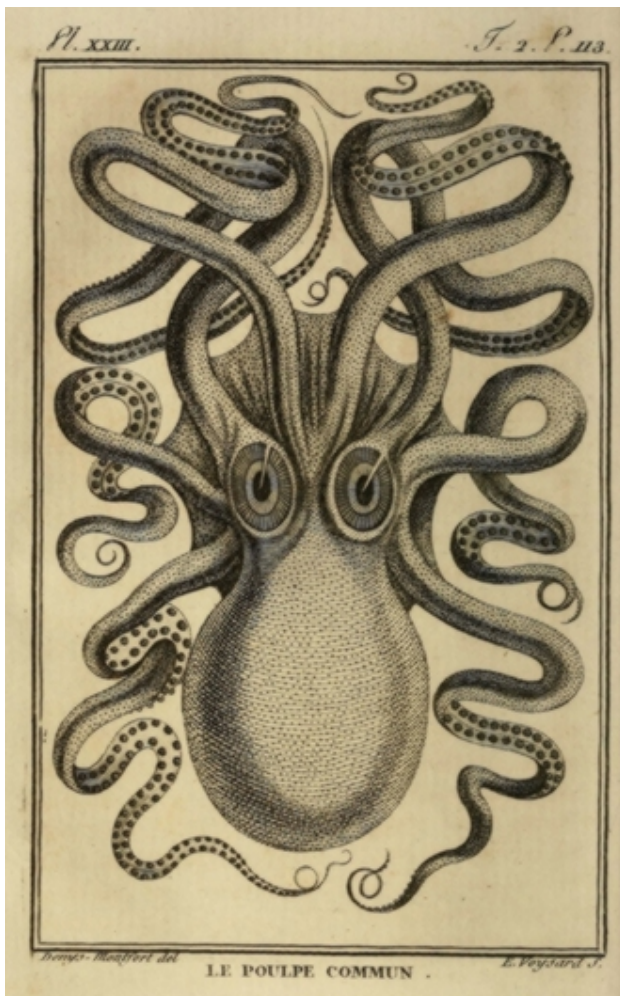
a. Les outils Google « Mon activité » (*My Activity*) <sup>7</sup> et « Téléchargez vos données » (*Takeout*) <sup>8</sup>, qui décrivent aux utilisateurs l'information collectée lors de l'usage

des outils Google.

b. Les données interceptées lors de l'envoi aux serveurs de Google pendant l'utilisation des produits Google ou d'autres sociétés associées.

c. Les règles concernant la vie privée (des produits Google spécifiquement ou en général).

d. Des recherches tierces qui ont analysé les collectes de données opérées par Google.



*Histoire naturelle, générale et particulière, des mollusques, animaux sans vertèbres et à sang blanc. T.2. Paris, L'Imprimerie de F. Dufart, An X-XIII [1802-1805].*  
[biodiversitylibrary.org/page/35755415](http://biodiversitylibrary.org/page/35755415)

4.  
Au travers de la combinaison des sources ci-dessus, cette étude montre une vue globale et exhaustive de l'approche de Google concernant la collecte des données

et aborde en profondeur certains types d'informations collectées auprès des utilisateurs et utilisatrices.

Cette étude met en avant les éléments clés suivants :

a. Dans une journée d'utilisation typique, Google en apprend énormément sur les intérêts personnels de ses utilisateurs. Dans ce scénario d'une journée « classique », où un utilisateur réel avec un compte Google et un téléphone Android (avec une nouvelle carte SIM) suit sa routine quotidienne, Google collecte des données tout au long des différentes activités, comme la localisation, les trajets empruntés, les articles achetés et la musique écoutée. De manière assez surprenante, Google collecte ou infère plus de deux tiers des informations via des techniques passives. Au bout du compte, Google a identifié les intérêts des utilisateurs avec une précision remarquable.

b. Android joue un rôle majeur dans la collecte des données pour Google, avec plus de 2 milliards d'utilisateurs actifs mensuels dans le monde <sup>9</sup>. Alors que le système d'exploitation Android est utilisé par des fabricants d'équipement d'origine (FEO) partout dans le monde, il est étroitement connecté à l'écosystème Google via le service Google Play. Android aide Google à récolter des informations personnelles sur les utilisateurs (nom, numéro de téléphone, date de naissance, code postal et dans beaucoup de cas le numéro de carte bancaire), les activités réalisées sur le téléphone (applications utilisées, sites web consultés) et les coordonnées de géolocalisation. En coulisses, Android envoie fréquemment la localisation de l'utilisateur ainsi que des informations sur l'appareil lui-même, comme sur l'utilisation des applications, les rapports de bugs, la configuration de l'appareil, les sauvegardes et différents identifiants relatifs à l'appareil.

c. Le navigateur Chrome aide Google à collecter des données utilisateurs depuis à la fois le téléphone et l'ordinateur de bureau, grâce à quelque 2 milliards d'installations dans le monde <sup>10</sup>. Le navigateur Chrome collecte des informations personnelles (comme lorsqu'un utilisateur remplit un formulaire en ligne) et les envoie à Google via le processus de synchronisation. Il liste aussi les pages visitées et envoie les données de géolocalisation à Google.

d. Android comme Chrome envoient des données à Google même en l'absence de toute interaction de l'utilisateur. Nos expériences montrent qu'un téléphone Android dormant et stationnaire (avec Chrome actif en arrière-plan) a

communiqué des informations de localisation à Google 340 fois pendant une période de 24 heures, soit en moyenne 14 communications de données par heure. En fait, les informations de localisation représentent 35 % de l'échantillon complet de données envoyés à Google. À l'opposé, une expérience similaire a montré que sur un appareil iOS d'Apple avec Safari (où ni Android ni Chrome n'étaient utilisés), Google ne pouvait pas collecter de données notables (localisation ou autres) en absence d'interaction de l'utilisateur avec l'appareil.

e. Une fois qu'un utilisateur ou une utilisatrice commence à interagir avec un téléphone Android (par exemple, se déplace, visite des pages web, utilise des applications), les communications passives vers les domaines de serveurs Google augmentent considérablement, même dans les cas où l'on n'a pas utilisé d'applications Google majeures (c.-à-d. ni recherche Google, ni YouTube, pas de Gmail ni Google Maps). Cette augmentation s'explique en grande partie par l'activité sur les données de l'éditeur et de l'annonceur de Google (Google Analytics, DoubleClick, AdWords) <sup>11</sup>. Ces données représentaient 46 % de l'ensemble des requêtes aux serveurs Google depuis le téléphone Android. Google a collecté la localisation à un taux 1,4 fois supérieur par rapport à l'expérience du téléphone fixe sans interaction avec l'utilisateur. En termes d'amplitude, les serveurs de Google ont communiqué 11,6 Mo de données par jour (ou 0,35 Go / mois) avec l'appareil Android. Cette expérience suggère que même si un utilisateur n'interagit avec aucune application phare de Google, Google est toujours en mesure de recueillir beaucoup d'informations par l'entremise de ses produits d'annonce et d'éditeur.

f. Si un utilisateur d'appareil sous iOS décide de renoncer à l'usage de tout produit Google (c'est-à-dire sans Android, ni Chrome, ni applications Google) et visite exclusivement des pages web non-Google, le nombre de fois où les données sont communiquées aux serveurs de Google demeure encore étonnamment élevé. Cette communication est menée exclusivement par des services de l'annonceur/éditeur. Le nombre d'appels de ces services Google à partir d'un appareil iOS est similaire à ceux passés par un appareil Android. Dans notre expérience, la quantité totale de données communiquées aux serveurs Google à partir d'un appareil iOS est environ la moitié de ce qui est envoyé à partir d'un appareil Android.

g. Les identificateurs publicitaires (qui sont censés être « anonymisés » et

collectent des données sur l'activité des applications et les visites des pages web tierces) peuvent être associés à l'identité d'un utilisateur ou utilisatrice de Google. Cela se produit par le transfert des informations d'identification depuis l'appareil Android vers les serveurs de Google. De même, le cookie ID DoubleClick (qui piste les activités des utilisateurs et utilisatrices sur les pages web d'un tiers) constitue un autre identificateur censé être anonymisé que Google peut associer à celui d'un compte personnel Google, si l'utilisateur accède à une application Google avec le navigateur déjà utilisé pour aller sur la page web externe. En définitive, nos conclusions sont que Google a la possibilité de connecter les données anonymes collectées par des moyens passifs avec les données personnelles de l'utilisateur.