

# Les données que récolte Google – Ch.7 et conclusion

*Voici déjà la traduction du septième chapitre et de la brève conclusion de [Google Data Collection](#), l'étude élaborée par l'équipe du professeur Douglas C. Schmidt, spécialiste des systèmes logiciels, chercheur et enseignant à l'[Université Vanderbilt](#). Si vous les avez manqués, retrouvez les [chapitres précédents déjà publiés](#).*

*Il s'agit cette fois-ci de tous les récents produits de Google (ou plutôt Alphabet) qui investissent nos pratiques et nos habitudes : des pages AMP aux fournisseurs de services tiers en passant par les assistants numériques, tout est prétexte à collecte de données directement ou non.*

*Traduction Framalang : Côme, Fabrice, goofy, Khrys, Piup, Serici*

## **VII. Des produits avec un haut potentiel futur d'agrégation de données**

83. Google a d'autres produits qui pourraient être adoptés par le marché et pourraient bientôt servir à la collecte de données, tels que AMP, Photos, Chromebook Assistant et Google Pay. Il faut ajouter à cela que Google est capable d'utiliser les données provenant de partenaires pour collecter les informations de l'utilisateur. La section suivante les décrit plus en détail.

84. Il existe également d'autres applications Google qui peuvent ne pas être largement utilisées. Toutefois, par souci d'exhaustivité, la collecte de données par leur intermédiaire

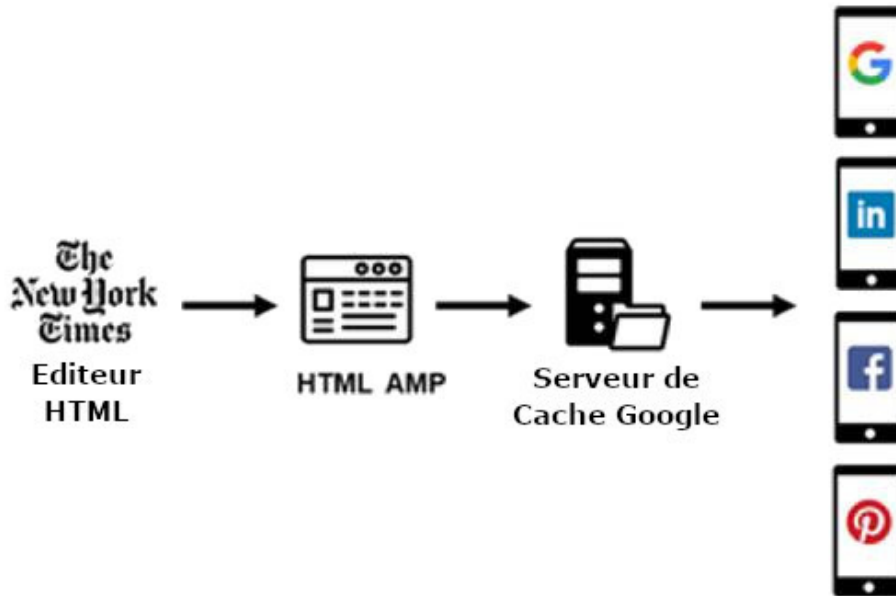
est présentée dans la section VIII.B de l'annexe.

## **A. Pages optimisées pour les mobiles (AMP)**

85. Les Pages optimisées pour les mobiles (AMP) sont une initiative *open source* menée par Google pour réduire le temps de chargement des sites Web et des publicités. AMP convertit le HTML standard et le code JavaScript en une version simplifiée développée par Google<sup>1</sup> qui stocke les pages validées dans un cache des serveurs du réseau Google pour un accès plus rapide<sup>2</sup>. AMP fournit des liens vers les pages grâce aux résultats de la recherche Google mais également via des applications tierces telles que LinkedIn et Twitter. D'après AMP : « L'écosystème AMP compte 25 millions de domaines, plus de 100 fournisseurs de technologie et plateformes de pointe qui couvrent les secteurs de la publication de contenu, les publicités, le commerce en ligne, les petits commerces, le commerce local etc. »<sup>3</sup>

86. L'illustration 17a décrit les étapes menant à la fourniture d'une page AMP accessible via la recherche Google. Merci de noter que le fournisseur de contenu à travers AMP n'a pas besoin de fournir ses propres caches serveur, car c'est quelque chose que Google fournit pour garantir un délai optimal de livraison aux utilisateurs. Dans la mesure où le cache AMP est hébergé sur les serveurs de Google, lors d'un clic sur un lien AMP produit par la recherche Google, le nom de domaine vient du domaine google.com et non pas du domaine du fournisseur. Ceci est montré grâce aux captures prises lors d'un exemple de recherche de mots clés dans l'illustration 17b.

a.



b.

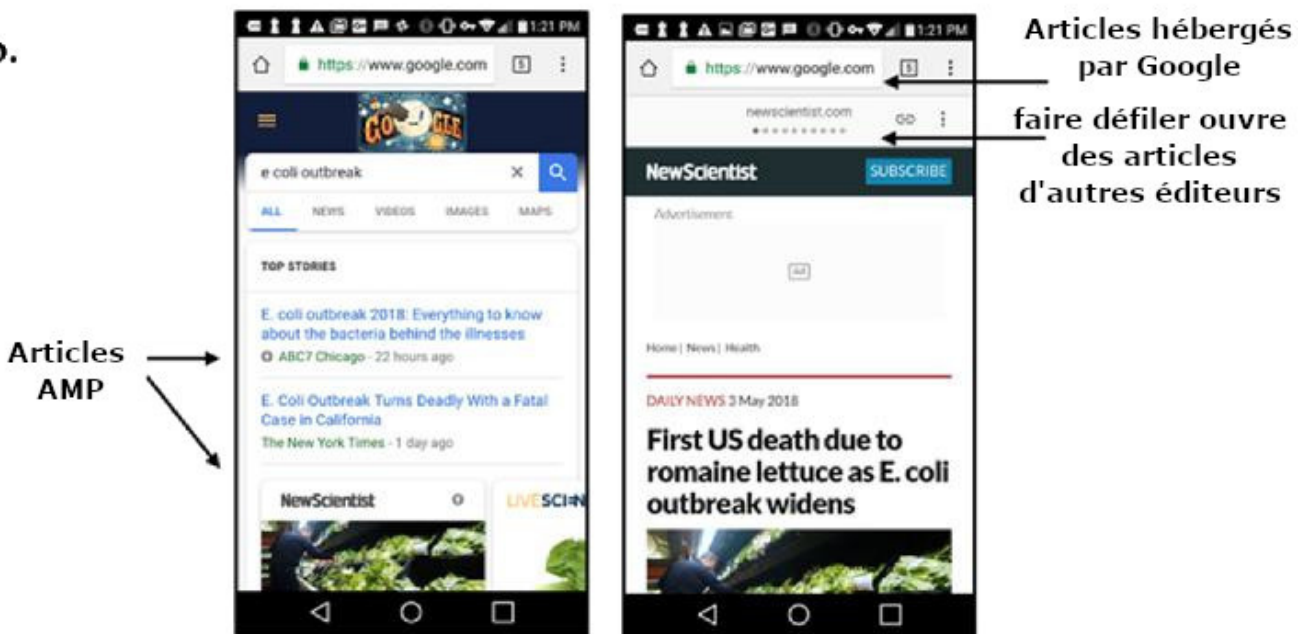


Illustration 17 : une page web normale qui devient une page AMP.

87. Les utilisateurs peuvent accéder au contenu depuis de multiples fournisseurs dont les articles apparaissent dans les résultats de recherche pendant qu'ils naviguent dans le carrousel AMP, tout en restant dans le domaine de Google. En effet, le cache AMP opère comme un réseau de distribution de contenu (RDC, ou CDN en anglais) appartenant à Google et géré par Google.

88. En créant un outil *open source*, complété avec un CDN, Google a attiré une large base d'utilisateurs à qui diffuser les sites mobiles et la publicité et cela constitue une quantité d'information significative (p.ex. le contenu lui-même, les pages vues, les publicités, et les informations de celui à qui ce contenu est fourni). Toutes ces informations sont disponibles pour Google parce qu'elles sont collectées sur les serveurs CDN de Google, fournissant ainsi à Google beaucoup plus de données que par tout autre moyen d'accès.

89. L'AMP est très centré sur l'utilisateur, c'est-à-dire qu'il offre aux utilisateurs une expérience de navigation beaucoup plus rapide et améliorée sans l'encombrement des fenêtres pop-up et des barres latérales. Bien que l'AMP représente un changement majeur dans la façon dont le contenu est mis en cache et transmis aux utilisateurs, la politique de confidentialité de Google associée à l'AMP est assez générale<sup>4</sup>. En particulier, Google est en mesure de recueillir des informations sur l'utilisation des pages Web (par exemple, les journaux de serveur et l'adresse IP) à partir des requêtes envoyées aux serveurs de cache AMP. De plus, les pages standards sont converties en AMP via l'utilisation des API AMP<sup>5</sup>. Google peut donc accéder à des applications ou à des sites Web (« clients API ») et utiliser toute information soumise par le biais de l'API conformément à ses politiques générales<sup>6</sup>.

90. Comme les pages Web ordinaires, les pages Web AMP pistent les données d'utilisation via Google Analytics et DoubleClick. En particulier, elles recueillent des informations sur les données de page (par exemple : domaine, chemin et titre de page), les données d'utilisateur (par exemple : ID client, fuseau horaire), les données de navigation (par exemple : ID et référence de page uniques), l'information du navigateur et les données sur les interactions et les événements<sup>7</sup>. Bien que les modes de collecte de données de Google n'aient pas changé

avec l'AMP, la quantité de données recueillies a augmenté puisque les visiteurs passent 35 % plus de temps sur le contenu Web qui se charge avec Google AMP que sur les pages mobiles standard.<sup>8</sup>

## **B. Google Assistant**

91. Google Assistant est un assistant personnel virtuel auquel on accède par le biais de téléphones mobiles et d'appareils dits intelligents. C'est un assistant virtuel populaire, comme Siri d'Apple, Alexa d'Amazon et Cortana de Microsoft.<sup>9</sup> Google Assistant est accessible via le bouton d'accueil des appareils mobiles sous Android 6.0 ou versions ultérieures. Il est également accessible via une application dédiée sur les appareils iOS<sup>10</sup>, ainsi que par l'intermédiaire de haut-parleurs intelligents, tel Google Home, qui offre de nombreuses fonctions telles que l'envoi de textes, la recherche de courriels, le contrôle de la musique, la recherche de photos, les réponses aux questions sur la météo ou la circulation, et le contrôle des appareils domestiques intelligents<sup>11</sup>.

92. Google collecte toutes les requêtes de Google Assistant, qu'elles soient audio ou saisies au clavier. Il collecte également l'emplacement où la requête a été effectuée. L'illustration 18 montre le contenu d'une requête enregistrée par Google. Outre son utilisation via les haut-parleurs de Google Home, Google Assistant est activé sur divers autres haut-parleurs produits par des tiers (par exemple, les casques sans fil de Bose). Au total, Google Assistant est disponible sur plus de 400 millions d'appareils<sup>12</sup>. Google peut collecter des données via l'ensemble de ces appareils puisque les requêtes de l'Assistant passent par les serveurs de Google.

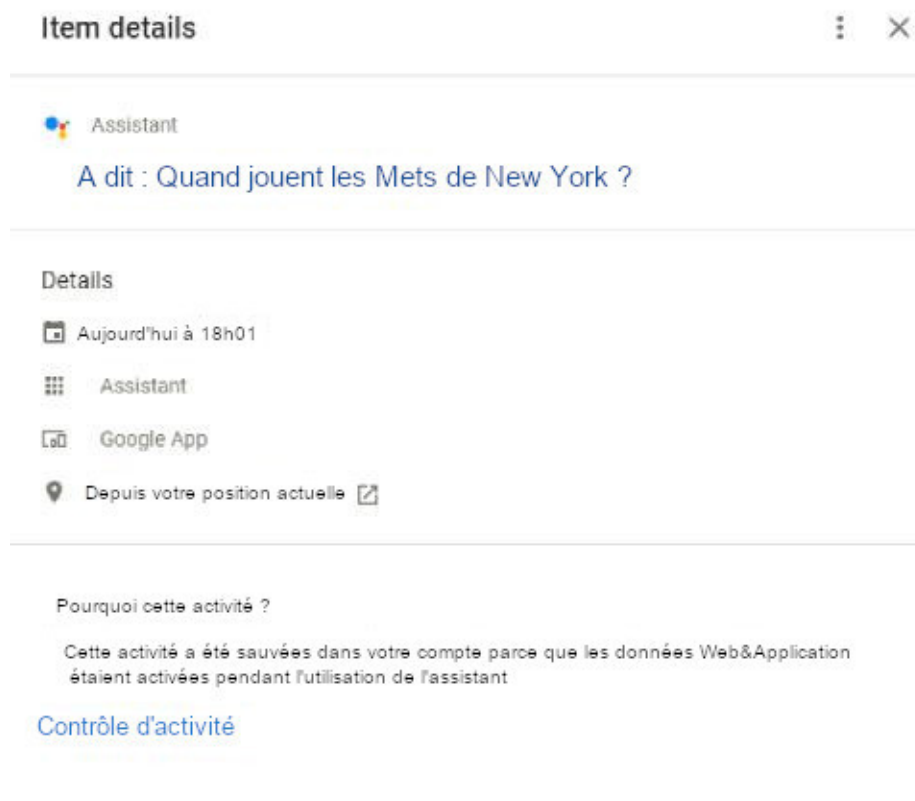


Figure 18 : Exemple de détails collectés à partir de la requête Google Assistant.

## C. Photos

93. Google Photos est utilisé par plus de 500 millions de personnes dans le monde et stocke plus de 1,2 milliard de photos et vidéos chaque jour<sup>13</sup>. Google enregistre l'heure et les coordonnées GPS de chaque photo prise. Google télécharge des images dans le Google cloud et effectue une analyse d'images pour identifier un large éventail d'objets, tels que les modes de transport, les animaux, les logos, les points de repère, le texte et les visages<sup>14</sup>. Les capacités de détection des visages de Google permettent même de détecter les états émotionnels associés aux visages dans les photos téléchargées et stockées dans leur cloud<sup>15</sup>.

94. Google Photos effectue cette analyse d'image par défaut lors de l'utilisation du produit, mais ne fera pas de distinction entre les personnes, sauf si l'utilisateur donne l'autorisation à l'application<sup>16</sup>. Si un utilisateur autorise Google à regrouper des visages similaires, Google identifie différentes personnes à l'aide de la technologie de reconnaissance faciale et permet aux utilisateurs de partager des photos grâce à sa technologie de « regroupement de visages »<sup>1718</sup>. Des exemples des capacités de classification d'images de Google avec et sans autorisation de regroupement des visages de l'utilisateur sont présentés dans l'illustration 19. Google utilise Photos pour assembler un vaste ensemble d'informations d'identifications faciales, qui a récemment fait l'objet de poursuites judiciaires<sup>19</sup> de la part de certains États.



Illustration : Exemple de reconnaissance d'images dans Google Photos.

## D. Chromebook

95. Chromebook est la tablette-ordinateur de Google qui fonctionne avec le système d'exploitation Chrome (Chrome OS)

et permet aux utilisateurs d'accéder aux applications sur le *cloud*. Bien que Chromebook ne détienne qu'une très faible part du marché des PC, il connaît une croissance rapide, en particulier dans le domaine des appareils informatiques pour la catégorie K-12, où il détenait 59,8 % du marché au deuxième trimestre 2017<sup>20</sup>. La collecte de données de Chromebook est similaire à celle du navigateur Google Chrome, qui est décrite dans la section II.A. Chromebooks permet également aux cookies de Google et de domaines tiers de pister l'activité de l'utilisateur, comme pour tout autre ordinateur portable ou PC.

96. De nombreuses écoles de la maternelle à la terminale utilisent des Chromebooks pour accéder aux produits Google via son service GSuite for Education. Google déclare que les données recueillies dans le cadre d'une telle utilisation ne sont pas utilisées à des fins de publicité ciblée<sup>21</sup>. Toutefois, les étudiants reçoivent des publicités s'ils utilisent des services supplémentaires (tels que YouTube ou Blogger) sur les Chromebooks fournis par leur établissement d'enseignement.

## **E. Google Pay**

97. Google Pay est un service de paiement numérique qui permet aux utilisateurs de stocker des informations de carte de crédit, de compte bancaire et de PayPal pour effectuer des paiements en magasin, sur des sites Web ou dans des applications utilisant Google Chrome ou un appareil Android connecté<sup>22</sup>. Pay est le moyen par lequel Google collecte les adresses et numéros de téléphone vérifiés des utilisateurs, car ils sont associés aux comptes de facturation. En plus des renseignements personnels, Pay recueille également des renseignements sur la transaction, comme la date et le montant de la transaction, l'emplacement et la description du marchand, le type de paiement utilisé, la description des articles achetés, toute photo qu'un utilisateur choisit



d'associer à la transaction, les noms et adresses électroniques du vendeur et de l'acheteur, la description du motif de la transaction par l'utilisateur et toute offre associée à la transaction<sup>23</sup>. Google traite ses informations comme des informations personnelles en fonction de sa politique générale de confidentialité. Par conséquent il peut utiliser ces informations sur tous ses produits et services pour fournir de la publicité très ciblée<sup>24</sup>. Les paramètres de confidentialité de Google l'autorisent par défaut à utiliser ces données collectées<sup>25</sup>.

## **F. Données d'utilisateurs collectées auprès de fournisseurs de données tiers**

98. Google collecte des données de tiers en plus des informations collectées directement à partir de leurs services et applications. Par exemple, en 2014, Google a annoncé qu'il commencerait à suivre les ventes dans les vrais commerces *réels* en achetant des données sur les transactions par carte de crédit et de débit. Ces données couvraient 70 % de toutes les opérations de crédit et de débit aux États-Unis<sup>26</sup>. Elles contenaient le nom de l'individu, ainsi que l'heure, le lieu et le montant de son achat<sup>27</sup>.

99. Les données de tiers sont également utilisées pour aider Google Pay, y compris les services de vérification, les informations résultant des transactions Google Pay chez les commerçants, les méthodes de paiement, l'identité des émetteurs de cartes, les informations concernant l'accès aux soldes du compte de paiement Google, les informations de facturation des opérateurs et transporteurs et les rapports des consommateurs<sup>28</sup>. Pour les vendeurs, Google peut obtenir des informations des organismes de crédit aux particuliers ou aux entreprises.

100. Bien que l'information des utilisateurs tiers que Google reçoit actuellement soit de portée limitée, elle a déjà attiré l'attention des autorités gouvernementales. Par exemple, la FTC a annoncé une injonction contre Google en juillet 2017 concernant la façon dont la collecte par Google de données sur les achats des consommateurs porte atteinte à la vie privée électronique<sup>29</sup>. L'injonction conteste l'affirmation de Google selon laquelle il peut protéger la vie privée des consommateurs tout au long du processus en utilisant son algorithme. Bien que d'autres mesures n'aient pas encore été prises, l'injonction de la FTC est un exemple des préoccupations du public quant à la quantité de données que Google recueille sur les consommateurs.

## VIII. CONCLUSION

101. Google compte un pourcentage important de la population mondiale parmi ses clients directs, avec de multiples produits en tête de leurs marchés mondiaux et de nombreux produits qui dépassent le milliard d'utilisateurs actifs par mois. Ces produits sont en mesure de recueillir des données sur les utilisateurs au moyen d'une variété de techniques qui peuvent être difficiles à comprendre pour un utilisateur moyen. Une grande partie de la collecte de données de Google a lieu lorsque l'utilisateur n'utilise *aucun* de ses produits directement. L'ampleur d'une telle collecte est considérable, en particulier sur les appareils mobiles Android. Et bien que ces informations soient généralement recueillies sans identifier un utilisateur unique, Google a la possibilité d'utiliser les données recueillies auprès d'autres sources pour *désanonymiser* une telle collecte.