



FICHE 3 : BIG DATA ET PUBLICITÉ SUR INTERNET

L'exploitation de données n'a jamais été aussi importante dans le secteur de la publicité en ligne. Dans l'écosystème de l'achat programmatique et des systèmes d'enchères en temps réel de nombreuses données sont exploitées, notamment pour améliorer le ciblage et les performances des campagnes. Les données qui sont exploitées proviennent de sources de plus en plus diverses. Les données sont générées par les individus, les entreprises, les pouvoirs publics, les machines... Les phénomènes sociaux (ex : circulation automobile) comme naturels (ex : météorologie) constituent également des sources de production de données, qui peuvent être utilisées de manière conjointe avec d'autres ensembles de données pour améliorer notamment le ciblage des publicités.

Les données elles-mêmes sont de plus en plus variées. Par exemple, les individus génèrent des données très différentes :

- données de contact (nom, prénom, e-mail...),
- données sociodémographiques,
- données de transaction,
- données relationnelles,
- données comportementales(recherches, types d'achats, publicités et vidéos visionnées...),
- données de géolocalisation,
- données de communication (messaging instantané, e-mails...),
- données techniques (identifiants d'appareils).

Ce développement concerne principalement les données non structurées¹ (fichiers audio, vidéos, fichiers textes, messages écrits, contenu des pages web...), qui représentent aujourd'hui la grande majorité des données produites².

Les données fournies et générées par les individus peuvent être fournies de manière active (ex : formulaire rempli) ou passive. Parmi les modes de collecte, il y a lieu de citer les formulaires³, les informations divulguées publiquement, les informations divulguées par des tiers (ex : commentaires sur des réseaux sociaux), les données obtenues par recoupement, autogénération et recherche automatisées ainsi que les traces collectées à l'insu de la personne (ex : cookies).

S'agissant des traces collectées à l'insu des personnes, plusieurs technologies sont utilisées. Les cookies constituent la technique la plus répandue sur les ordinateurs. Il

¹ Contrairement aux données structurées, elles ne sont pas organisées selon un modèle prédéfini (ex : formats des données et relations entre les données).

² OCDE, Data-Driven Innovation, Big Data for Growth and Well-Being, octobre 2015, p.151.

³ Formulaires d'abonnement (service, revue etc), données de transaction lors d'un achat en ligne (adresse de livraison, données bancaires, âge, genre, préférences, etc.).

s'agit de fichier texte déposé sur le disque dur d'une personne par le serveur web d'un site visité ou par le serveur web utilisé par un acteur tiers, notamment actif dans le secteur de la publicité. Il existe de nombreuses autres techniques parmi lesquelles :

- Les empreintes numériques, qui permettent de reconnaître une personne ou un appareil, à partir de la collecte et de la combinaison d'un ensemble de données qui caractérisent un appareil,
- Les pixels, qui sont des petites images numériques placées sur un site web et qui permettent notamment de collecter des habitudes de navigation,
- Le GPS et les réseaux mobiles peuvent être utilisés par les opérateurs de communications électroniques pour obtenir des données de localisation,
- Les bornes Wi-Fi présentes dans certains magasins peuvent être utilisées à cette fin par les commerçants,
- Les identifiants publicitaires (Advertising ID) présents sur les terminaux mobiles opérant sous les systèmes d'exploitation de Google ou d'Apple, qui permettent de transmettre et de collecter des données de navigation ,de téléchargement ou de navigation, et plus largement toute activité réalisée dans le cadre du système d'exploitation.