

Les données de la publicité numérique : de l'accumulation à la redéfinition de la valeur et des pratiques

Par Théophile MEGALI

Chaire Gouvernance et Régulation de l'Université Paris-Dauphine_PSL

Le fonctionnement de la publicité numérique repose sur l'utilisation massive de données, principalement pour l'achat/vente des espaces, le ciblage ou la mesure de l'audience et de la performance des campagnes. De par leur intégration verticale, les plateformes fonctionnant en « jardins clos » suivent une logique d'accumulation de ces différentes données, et se distinguent de leurs concurrents. Toutefois, à ces éléments structurels s'ajoutent des facteurs juridiques et techniques qui viennent provoquer une redéfinition de la valeur des données et une évolution des pratiques publicitaires, à nouveau dans un sens favorable à ces acteurs.

En publicité comme dans l'ensemble du monde numérique, il est désormais bien connu que nos traces valent autant que notre présence, nos données autant que notre attention : les deux sont mobilisées de pair par les acteurs publicitaires. À la valeur attribuée à l'attention des internautes, qu'il s'agisse du temps passé sur un contenu ou de l'interaction réalisée avec celui-ci, s'ajoute désormais la valorisation des données issues des caractéristiques et préférences de ces internautes. Ces données sont cruciales à l'opérationnalisation du ciblage, qui constitue l'un des traits saillants de la publicité en ligne. Cette capacité de ciblage contribue à garantir aux supports numériques une attractivité auprès des marques qui augmente d'année en année : en France, les recettes nettes de la publicité sur Internet ont augmenté de 29 % en 2021 par rapport à 2019, tandis que la taille de ce marché a plus que doublé (+ 120 %) entre 2016 et 2021¹.

Ce dynamisme profite principalement à un triopole d'acteurs (Google, Facebook et Amazon) qui a capté 67 % du total des recettes nettes en France en 2021², en se répartissant sur différents segments (dont, principalement, les requêtes sponsorisées sur les moteurs de recherche (*search advertising*) et la publicité graphique sur bannières et vidéos (*display*) sur les sites *web*, applications et réseaux sociaux). Les recettes restantes (33 %) sont partagées entre une multiplicité d'acteurs venus notamment des médias, du secteur de la distribution ou opérés complètement en ligne (*pure players*).

Cette asymétrie de revenus, entre un triopole très puissant et une frange de concurrents, se matérialise également pour ce qui concerne l'échange et l'utilisation des données à des fins publicitaires. Tandis que la capacité d'extraction et de mobilisation des données personnelles par les acteurs du marché publicitaire tend à s'amenuiser, les plateformes semblent moins sensibles à cette évolution, voire de plus en plus en mesure d'arbitrer la définition de la valeur de ces données.

¹ SRI, UDECAM & Oliver Wyman, *27^e observatoire de l'e-pub*, Bilan FY 2021, 2022.

² SRI *et al.*, *ibid.*

DES DONNÉES INDISPENSABLES À LA MÉCANIQUE PUBLICITAIRE SUR INTERNET

La publicité numérique est largement automatisée dans ses processus d'achat/vente et d'allocation des espaces : ce phénomène est désigné sous le terme de « publicité programmatique » et concerne la publicité *search*, la publicité sur les réseaux sociaux et une large partie (64 % en 2021³) de la publicité *display*. Ce mode s'oppose à la vente de gré à gré, omniprésente dans les médias traditionnels et désormais marginale pour la publicité numérique. L'organisation de la publicité « programmatique » repose directement sur l'usage de données, par plusieurs aspects principaux.

- Un premier aspect concerne l'utilisation de données pour faire fonctionner l'enchère et l'allocation dynamique des publicités. La fixation des prix des espaces ou des mots-clés (dans le cas des requêtes sponsorisées) repose sur des algorithmes mobilisant notamment des informations sur les propriétés du profil de l'utilisateur qui sera exposé à cette publicité, sur le prix des campagnes passées ou sur le niveau de demande pour cet espace publicitaire. Mais ce processus peut parfois tenir compte de la pertinence de la campagne, fondée notamment sur l'adéquation entre les attentes des utilisateurs et les caractéristiques de la campagne de l'annonceur⁴.
- Un deuxième aspect concerne les *analytics*, ces outils de mesure de la performance et de l'audience des espaces publicitaires. Les *analytics* constituent, de manière générale, un retour d'information relatif au comportement des internautes sur un site (mesure *site-centric*) ou à l'égard d'une publicité (mesure *ad-centric*). Dans cette dernière catégorie, les données produites sont relatives aux actions réalisées vis-à-vis des publicités : nombre de fois où elles ont été vues (impressions) ou cliquées, ou nombre de fois où elles ont mené à un achat (conversion), par exemple. Ces données sont indispensables pour évaluer la performance des campagnes et sont collectées par les outils techniques intermédiaires responsables de l'achat média pour le compte de l'annonceur (*demand side platforms*) ou de la diffusion des publicités (*ad servers*) (Beneteau, 2018).
- Un dernier aspect voit l'utilisation des données à des fins de ciblage de l'audience et de personnalisation. Lors des enchères, la mobilisation de données (personnelles) sur l'utilisateur qui consulte le site ou l'application sur laquelle un espace est à vendre permet de mieux valoriser celui-ci. Les données mobilisées sont collectées par les annonceurs ou les éditeurs eux-mêmes (dites *first party*), ou par d'autres annonceurs (*second party*), ou encore par des tiers spécialisés dans la collecte et l'exploitation de données (*third party*). Elles dressent le profil socio-démographique de l'internaute ou permettent de révéler, à la suite de déclarations ou par inférence, ses préférences ou ses intentions d'achat. Par ce biais, les annonceurs sont en mesure de cibler des bassins d'audience à travers le *web* et sans se soucier du contexte éditorial dans lequel sera diffusée la publicité. Ce ciblage permet alors d'adapter le contenu des messages et de le personnaliser de manière dynamique, en fonction des caractéristiques attribuées à l'internaute (par exemple rabais personnalisé, reciblage d'un produit déjà consulté), dans une logique de performance pour adresser « le bon message, à la bonne personne et au bon moment » (Crain, 2019).

³ SRI *et al.*, *ibid.*

⁴ Cela vaut notamment pour la publicité *search* sur Google qui tient compte, dans l'attribution des enchères, d'un *quality score* (<https://support.google.com/google-ads/answer/6167118?hl=fr>), ou de Facebook qui mesure également la qualité des publicités (<https://www.facebook.com/business/help/430291176997542?id=561906377587030>)

UN MODÈLE PUBLICITAIRE FONDÉ SUR L'ACCUMULATION DES DONNÉES

L'économie numérique a donc grandement bénéficié de la baisse des coûts liés au ciblage, et a permis l'amélioration de la discrimination dans les prix – notamment en matière publicitaire comme nous venons de le voir – et l'émergence d'un ciblage particulièrement fin (Goldfarb et Tucker, 2019).

Il convient cependant de souligner que, si toute l'économie de la publicité dite « programmatique » repose sur l'exploitation de ces données, les acteurs de ce secteur se rendent donc dépendants de ceux qui les collectent. Le travail de valorisation – et ses bénéfices – échoit principalement aux acteurs marchands : la constitution de bases de données, leur entretien permanent (en raison, notamment, de la « péremption » des données utiles au ciblage) constituent effectivement l'effort nécessaire au fonctionnement du marché publicitaire. Ces opérations contribuent à la perception de données comme « actifs » (Beauvisage et Mellet, 2020) valorisables par les acteurs marchands, tout en faisant, par la même occasion, émerger un enjeu autour de la qualité de ces bases de données (Megali et Quinn, 2021).

La nécessité de pouvoir disposer de bases d'une taille suffisante, pour le ciblage notamment, a incité certains acteurs à trouver de nouvelles formes de gouvernance autour des données (par exemple de l'alliance Gravity, réunissant 250 sites français) ou à trouver de nouveaux débouchés (par exemple des fournisseurs d'accès à Internet qui valorisent désormais une partie des données de consommation TV de leurs clients auprès des éditeurs TV dans le cadre d'offres de publicité TV segmentée). Ces démarches s'inscrivent toutefois dans un paysage éclaté, tant en termes d'offres de données que d'intermédiaires techniques permettant d'opérer des campagnes ciblées dont l'efficacité est de plus en plus contestée (Hwang, 2022).

Face à cela, les plateformes en « jardins clos » tirent leur épingle du jeu. Google et Facebook, par exemple, proposent un achat automatisé contrôlé de bout en bout et leur permettant de maîtriser les enchères, l'adressage technique, le ciblage et, enfin, la mesure de l'audience et de l'efficacité des campagnes. Ces solutions intégrées ont l'avantage de la simplicité de pilotage pour les annonceurs, notamment ceux qui n'ont pas les moyens d'avoir recours aux services d'agences média. Par ce biais, les plateformes développent un environnement technique où les données sont produites et valorisées de manière intégrée, dans une logique d'accumulation. Par ailleurs, le modèle des « jardins clos » repose également sur des services au grand public utilisant des identifiants uniques permettant de maximiser la collecte de données. L'ensemble des traces laissées par les utilisateurs à l'occasion de leurs interactions sur ces services contribue à générer des volumes colossaux de données utilisées pour catégoriser les utilisateurs et offrir une capacité de profilage inégalable aux annonceurs.

Deux modèles s'opposent donc sur le marché publicitaire numérique. D'un côté, un ensemble large d'acteurs fonctionnant dans des systèmes ouverts, qui tentent d'accroître la valorisation tirée des données qu'ils coproduisent et exploitent. Cette manière de procéder implique nécessairement un partage de la valeur entre producteurs de données, intermédiaires techniques et autres acteurs tiers permettant d'aboutir à la mise en œuvre de campagnes. De l'autre, des entreprises intégrées verticalement, fonctionnant en silo et de manière autosuffisante, capables de gérer à elles seules l'intégralité des campagnes publicitaires. Pour ces plateformes, l'accumulation de données permet d'entraîner l'ensemble des algorithmes des services offerts aux consommateurs et aux publicitaires, d'augmenter leur précision et leur efficacité, et donc de renforcer leurs parts de marché. L'augmentation de l'utilisation de ces services renforce leur attractivité par la présence d'externalités de réseau, qui viennent doper l'attractivité des services publicitaires et maximiser les revenus générés.

VERS UNE REDÉFINITION DE LA VALEUR DES DONNÉES PUBLICITAIRES

La démonstration du succès des plateformes pourrait se limiter à l'explication de ces mécanismes faisant de l'intégration verticale un outil puissant sur le marché publicitaire. Toutefois, d'autres facteurs juridiques et techniques viennent s'ajouter à ces éléments structurels et contribuent, par effet de bord, à renforcer la prédominance de ces plateformes.

Des évolutions juridiques, au premier titre desquelles la mise en œuvre du Règlement général sur la protection des données (RGPD)⁵, qui sont venues renforcer la transparence des usages et le contrôle des utilisateurs sur le recueil des données les concernant et, partant, sevrer une partie du marché publicitaire de l'utilisation abusive et incontrôlée des données personnelles. Une étude de Peukert, Bechtold, Batikas & Kretschmer (2022) a permis de démontrer que l'entrée en vigueur du RGPD a abouti à une baisse des échanges des sites *web* avec des tierces parties collectant des données. En outre, cette baisse n'a pas été également répartie, et une réduction moindre des capacités de collecte a donné lieu à certaines entreprises, au premier titre desquelles Google, de gagner en parts de marché sur certains segments (tels que celui des *analytics*).

Le processus de mise en conformité des acteurs publicitaires à ce texte européen apparaît malaisé, dans la mesure où la collecte de données demeure le nerf de la guerre sur le marché publicitaire. En France, plusieurs décisions récentes de la CNIL (Commission nationale de l'informatique et des libertés) sont ainsi intervenues pour venir sanctionner les acteurs ne permettant pas aux internautes de refuser aussi facilement que d'accepter la collecte de leurs données personnelles⁶. De nombreux éditeurs de sites cherchent effectivement à ruser pour inciter les internautes à accorder leur consentement par le biais de dispositifs trompeurs (*dark patterns*), mais cette pratique devrait connaître une limitation par voie réglementaire avec le "Digital Services Act"⁷. Progressivement, ce cadre juridique est venu amoindrir les capacités de collecte et rappeler qu'une « bonne » donnée est une donnée collectée après recueil du consentement.

À cet aspect juridique s'ajoute un changement au niveau industriel. Depuis plusieurs années, les contrôleurs d'accès (*gatekeepers*) du *web* sur ordinateur et sur mobile sont venus restreindre fortement la collecte de données sur des sites tiers. Dans la suite de Safari (Apple) et de Firefox (Mozilla), Chrome (Google) a ainsi annoncé la fin des *cookies* tiers sur son navigateur⁸. De même, les deux systèmes d'exploitation d'Apple et de Google vont limiter l'accès aux identifiants publicitaires uniques de leurs utilisateurs⁹. Face au bouleversement créé par cette restriction à la collecte de données, Google a lancé deux

⁵ Règlement 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

⁶ Voir notamment les sanctions récentes à l'égard de Google et de Facebook (délibération de la formation restreinte n°SAN-2021-023 et n°SAN-2021-024), ainsi que la mise en demeure d'une soixantaine d'autres acteurs en décembre 2021 : <https://www.cnil.fr/fr/refuser-cookies-doit-etre-aussi-simple-accepter-cnil-poursuit-action-adresse-mises-en-demeure>

⁷ Parlement européen, communiqué de presse : "Digital Services Act: regulating platforms for a safer online space for users", 20 janvier 2022.

⁸ Chromium Blog, "Justin Schuh - Building a more private web: A path towards making third party cookies obsolete", 14 janvier 2020 : <https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html>

⁹ Alice Vitard, « Avec la Privacy Sandbox, Google veut limiter le suivi publicitaire sur les applications Android », *usine-digitale.fr*, 17 février 2022.

initiatives : Privacy Sandbox¹⁰ et Topics API¹¹ permettant d'aboutir à des ciblage par cohortes d'individus ou à restreindre la connaissance des préférences individuelles à un nombre limité d'intérêts du moment. D'autres acteurs du marché font par ailleurs émerger des solutions alternatives offrant de miser davantage sur le contexte éditorial, par l'adéquation entre le contenu publicitaire et celui de la page visitée. En ce sens, certains types de données, suivant leur mode de collecte, sont amenés à disparaître ou à être fortement dévalorisés.

Ces changements vont faire évoluer la création de valeur sur le marché publicitaire, quitte à susciter des interrogations du point de vue concurrentiel. L'extinction progressive des *cookies* tiers ou la restriction à l'accès aux identifiants publicitaires sur mobile devrait faire des environnements « identifiés » des lieux où la connaissance des utilisateurs et de leur comportement (à des fins de ciblage, de personnalisation ou de mesure de l'audience) demeurera privilégiée. Cela pourrait ainsi participer du renforcement de la compétitivité du modèle des plateformes fonctionnant en « jardin clos » ou, de façon plus générale, des acteurs collectant sur la base d'un identifiant unique.

La restriction au ciblage tous azimuts oblige à repenser plus généralement la manière d'opérer la monétisation des sites (*via* les partenariats noués entre éditeurs et acteurs tiers) et le ciblage publicitaire à la maille individuelle, à l'heure où les enjeux de *privacy* sont de plus en plus prégnants. Certains acteurs, tels que les plateformes, conservent toutefois une capacité technique permettant un ciblage très précis, et s'inscrivent toujours dans une logique d'accumulation. Ils sont cependant ceux qui participent directement, voire impulsent la redéfinition collective de la valeur des données et des pratiques publicitaires.

RÉFÉRENCES BIBLIOGRAPHIQUES

BEAUVISAGE T. & MELLET K. (2020), "Datassets: Assetizing and marketizing personal data", in BIRCH K. & MUNIESA F. (éd.), *Assetization : Turning Things into Assets in Technoscientific Capitalism*, MIT Press, pp. 75-95.

BENETEAU P.-E. (2018), « Chapitre 1. Collecte, identification et exploitation des données », in ALLARY J. & BALUSSEAU V. (éd.), *La publicité à l'heure de la data*, Dunod, pp. 25-68.

CRAIN M. (2019), "A critical political economy of web advertising history", in BRÜGGER N. & MILIGAN I. (éd.), *The SAGE Handbook of Web History*, SAGE London, pp. 397-410.

GOLDFARBA. & TUCKER C. (2019), "Digital economics", *Journal of Economic Literature*, 57(1), pp. 3-43.

HWANG T. (2022), *Le grand krach de l'attention. La publicité, une bombe au cœur de l'internet*, C&F Éditions.

MEGALI T. & QUINN M. (2021), "Does quality really matter? Exploring data quality assessment in the French online advertising market", *Working Paper de la Chaire Gouvernance et Régulation*.

PEUKERT C., BECHTOLD, S., BATIKAS M. & KRETSCHMER T. (2022), "Regulatory spillovers and data governance: Evidence from the GDPR", *Marketing Science*.

¹⁰ Chetna Bindra, « Mieux comprendre la Privacy Sandbox, l'alternative de Chrome aux cookies-tiers », *Blog Think with Google*, septembre 2021 : <https://www.thinkwithgoogle.com/intl/fr-fr/futur-du-marketing/confidentialite-et-confiance/comprendre-privacy-sandbox/>

¹¹ Emma Roth, "Google abandons FLoC, introduces Topics API to replace tracking cookies", *The Verge.com*, 25 janvier 2022.