

Mémoire de Master
Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales
(MIASHS)
Parcours Web-Analyste

Un monde sans cookies

Le monde sans cookies qui se prépare entraîne-t-il la fin de la collecte
de données ?

Rédigé par Guillaume Sinnaeve
Janvier 2021

Mémoire encadré par Charles Paperman, maître de conférences à l'université de Lille.

Remerciements

Je voudrais dans un premier temps remercier mon directeur de mémoire monsieur Charles Paperman, responsable pédagogique et maître de conférence en informatique à l'université de Lille, pour sa disponibilité, ses judicieux conseils et sa passion pour le code que je partage et qui ont contribué à alimenter ma réflexion sur le projet et le mémoire.

Je remercie également madame Carine Wos, conseillère en formation continue à l'université de Lille, pour sa réactivité dans nos échanges et son accompagnement dans mon inscription au master Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales (MIASHS) parcours Web Analyse par la Validation des Acquis de l'Expérience (VAE).

Je tiens à témoigner toute ma reconnaissance aux personnes suivantes, pour leur aide dans la réalisation de ce mémoire :

Monsieur Maxence Gama, Chef de projet web analyse chez OUI.sncf, pour nos discussions et nos débats sur la finalité de la collecte de données qui ont contribué à ma réflexion.

Monsieur Guillaume Tollet, Consulting Associate Director et DPO chez fifty-five, pour sa compréhension et ses explications claires et précises sur les textes du RGPD et de la CNIL.

Monsieur François Khoury, Senior Manager et expert sur la web analyse, le mobile et le CRM chez fifty-five, qui m'a partagé ses connaissances et ses expériences dans le domaine de la collecte de données.

À mes interlocuteurs au sein des entreprises qui ont souhaité garder l'anonymat pour avoir répondu à mes questions à propos des impacts des textes législatifs sur la collecte de données, ainsi que sur leur expérience personnelle. Les questionnaires sont disponibles en annexe de ce mémoire.

Madame Marjorie Braconnier, Tools Senior Manager chez fifty-five, et Monsieur Clément Somon, Senior Tracking Specialist chez fifty-five, pour avoir relu et corrigé mon mémoire. Leurs conseils de rédaction ont été très précieux.

Ma femme, pour son soutien constant et ses encouragements.

Sommaire

Remerciements	1
Sommaire	2
Introduction	4
Partie 1 - La disparition des cookies	6
1.1 La finalité de la collecte de données	6
1.1.1 Mesurer son audience	7
1.1.2 Optimiser la conversion	8
1.1.3 Monétiser son contenu	8
1.2 L'émergence du cookie	9
1.2.1 La genèse	9
1.2.2 La donnée personnelle en péril	10
1.2.3 La croissance d'un modèle économique	11
1.3 La fin des cookies et ses conséquences	13
1.3.1 Le RGPD et la directive ePrivacy dictent les règles du jeu	13
1.3.2 Apple et la confidentialité des utilisateurs	14
1.3.3 Firefox et le consentement au niveau du navigateur	15
1.3.4 Google et la promesse du respect de l'anonymat	17
1.3.5 Brave et la protection de la vie privée	18
1.3.6 Les extensions bloquent la publicité	19
1.4 Un enjeu de taille pour les entreprises	20
1.4.1 Une connaissance client diminuée	20
1.4.2 Une organisation juridique	21
1.4.3 Un investissement technique	21
Partie 2 - Concilier respect du RGPD et exhaustivité de la connaissance utilisateur	22
2.1 Exemption CNIL	23
2.1.1 Collecte de données dès la première page	23
2.1.2 Une finalité limitée	23
2.2 Server-side	24
2.2.1 Réduire la charge client pour un meilleur contrôle	24
2.2.2 Consentement et temps de développement	25
2.3 Navigation authentifiée	25
2.3.1 Une relation personnalisée	26
2.3.2 Frein potentiel à la conversion de l'utilisateur	26
2.4 Modélisation et Machine Learning	27
2.4.1 Construire des données	27
2.4.2 Estimation approximative	28
2.5 Un consentement souvent nécessaire	28

Partie 3 - Expérimentation d'une collecte de données sans cookies, sans JavaScript et conforme au RGPD	29
3.1 Les pré-requis législatifs	29
3.1.1 Une finalité stricte	29
3.1.2 Informer les utilisateurs	31
3.2 Les contraintes techniques	32
3.2.1 Collecte des données en PHP	33
3.2.2 Mesure des interactions en CSS	34
3.2.3 Identifiant visiteur en paramètre d'URL	36
3.3 De la collecte à la visualisation des données	37
3.3.1 Connaître ses visiteurs	37
3.3.2 Le comportement sur le site	39
3.3.3 Traitement et stockage	40
3.3.4 Représentation graphique des données	43
Conclusion	47
Annexes	49
Annexe 1 - Questionnaire - Impact du RGPD sur Entreprise A	49
Annexe 2 - Questionnaire - Impact du RGPD sur Entreprise B	53
Annexe 3 - Questionnaire - Impact du RGPD sur Entreprise C	57
Annexe 4 - Questionnaire - Impact du RGPD sur Entreprise D	61
Annexe 5 - Politique de confidentialité	64
Annexe 6 - Événement de suppressions des données de plus de 25 mois	65
Annexe 7 - Campagnes d'acquisition de données	66
Annexe 8 - Schéma de la base de données	71
Annexe 9 - Page d'analyse du site	73
Table des illustrations	77
Webographie	79

Introduction

“La data c’est mon dada”. J’aime beaucoup cette phrase car, en plus d’être drôle à dire, elle exprime bien mon quotidien. Depuis 2015, la donnée est devenue mon métier. Tout ce qui nous entoure est de la donnée. Nos sens captent constamment de l’information. Un ciel bleu, le bruit des vagues, les grains de sable sous nos pieds. En associant ces quelques mots, c’est la plage et la mer qui nous viennent à l’esprit, un souvenir de vacances en été.

Aujourd’hui, la donnée est essentielle pour les entreprises. La donnée leur permet de connaître leurs utilisateurs et leurs demandes, de comprendre les parcours des visiteurs sur leur site internet et application mobile ou encore les revenus générés par leurs services. Ces entreprises rêvent d’un monde où elles pourraient avoir des données sur tout et sur tous pour mieux comprendre et anticiper les besoins de chacun.

Est-ce vraiment raisonnable et utile de savoir tout de tout le monde me direz vous ? Non. Imaginez à quel point une fuite de données personnelles peut être problématique pour les personnes visées. Comme ce fut le cas pour l’application de rencontre gay Grindr en avril 2018.

“L’application de rencontres homosexuelles a laissé des entreprises tierces accéder à des données privées de ses utilisateurs, dont leur statut VIH.”

Source : Le Monde

C’est pourquoi il est crucial de protéger la vie privée et les libertés individuelles. En Europe, le Règlement Général sur la Protection des Données (RGPD en français et GDPR en anglais pour General Data Protection Regulation) a bouleversé profondément la collecte de la donnée depuis mai 2018, en s’attaquant de front au problème des données personnelles. Les entreprises doivent se mettre en conformité avec ces textes de loi. Dans le cas contraire, elles s’exposent à des sanctions pouvant aller jusqu’à 20 millions d’euros ou 4% du chiffre d’affaires. Carrefour France a été contrôlé et condamné par la CNIL en novembre 2020.

“RGPD : la CNIL condamne Carrefour à plus de 3 millions d’euros d’amendes”

Source : La Tribune

En complément du RGPD, les entreprises se doivent de respecter la directive ePrivacy¹, autrement appelée “directive vie privée et communication électroniques”. Elle concerne le traitement des données à caractère personnel et la protection de la vie privée. La directive ePrivacy s’applique au secteur des communications électroniques et notamment aux sites Web qui doivent demander le consentement de l’utilisateur avant de collecter la donnée. Cependant,

¹ Directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (directive vie privée et communications électroniques) : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32002L0058>

elle n'a pas été mise à jour depuis juillet 2002 et peut créer des différences d'interprétation avec le RGPD de 2018.

Chaque Etat membre de l'Union Européenne dispose d'un organisme qui aide les professionnels à se mettre en conformité avec la législation. En France, c'est la Commission Nationale de l'Informatique et des Libertés (CNIL) qui est le régulateur des données personnelles. Les séances plénières de la CNIL, du 17 octobre 2019 et plus récemment celle de Septembre 2020, appuyées par le Conseil d'Etat, affirment les positions de la France sur la directive ePrivacy et la demande de consentement.

“En application de la directive ePrivacy, les internautes doivent être informés et donner leur consentement préalablement au dépôt et à la lecture de certains traceurs, tandis que d'autres sont dispensés du recueil de ce consentement.”

Source : CNIL

Les principaux acteurs et les entreprises réagissent et semblent aller dans la même direction avec réserve. Cependant, c'est toute l'industrie de la collecte de données qui est mise en péril. L'utilisation du cookie est bien maîtrisée depuis plusieurs années et des solutions de contournement face aux navigateurs et aux bloqueurs de publicité ont été trouvées avec le temps. Mais jusqu'à quand ? Quel futur se dessine pour cette industrie si bien organisée ? Le monde qui se prépare entraîne-t-il la fin de la collecte de données ? L'objectif de ce mémoire est de montrer que nous assistons à une transformation des méthodes de collecte de données et que l'ère des cookies telle que nous la connaissons touche à sa fin.

Dans une première partie, nous allons montrer la place importante qu'a pris le cookie dans notre quotidien sur le Web. Nous comprendrons pourquoi les acteurs législatifs et technologiques actuels ainsi que la prise de conscience des utilisateurs sur l'intérêt à protéger leurs informations personnelles mènent la vie dure aux cookies, aux solutions de collecte et aux entreprises pilotées par la donnée.

Ensuite, nous étudierons diverses solutions existantes qui permettent d'accorder le respect et la conformité au RGPD avec une connaissance la plus complète possible sur les utilisateurs. Comprendre comment les avantages et les inconvénients de chaque solution présentée permettent aux entreprises de faire un choix sur celle qui serait la plus adaptée à leur environnement. Nous concluons en montrant que la solution parfaite n'existe pas aujourd'hui.

Finalement, nous expérimenterons la mise en place d'une collecte de données côté serveur, sans cookies, sans JavaScript et conforme au RGPD et à la CNIL sur le site internet www.guillaume-sinnaeve.com. Nous aborderons les choix techniques mis en place pour respecter les règles énoncées précédemment, la mécanique utilisée pour mesurer les informations sur les pages et interactions des visiteurs et la visualisation graphique des données collectées.

Partie 1 - La disparition des cookies

Dans le domaine de la collecte de données sur le Web, que ça soit à des fins statistiques ou publicitaires, les cookies sont omniprésents. Un modèle économique entier repose sur les cookies et les informations qu'ils contiennent. Mais face aux acteurs législatifs et technologiques actuels, la fin des cookies semble inévitable et ne sera pas sans répercussions pour les entreprises pilotées par la donnée.

1.1 La finalité de la collecte de données

Collecter de la donnée utilisateur sur un site internet est nécessaire pour l'entreprise. Cependant, les utilisateurs qui font l'objet d'un tel suivi ne sont pas tous en mesure d'en comprendre les objectifs. Ainsi, la mise en place d'un outil de gestion du consentement des utilisateurs nécessite de définir l'objectif de collecte d'une donnée et de chaque technologie de suivi. C'est ce que le marché appelle la définition des "finalités". Si une entreprise ne sait pas répondre à la question "pourquoi nous collectons cette information ?" alors il n'est pas pertinent de la collecter.

Les cookies et autres traceurs sont réunis autour de grandes familles avec des finalités. Celles-ci répondent à la question de l'utilité des familles de cookies pour lesquelles la demande de consentement est faite à l'utilisateur. Ce classement des cookies par finalité est normé, mais des changements peuvent tout de même être faits et adaptés pour chaque entreprise. La CNIL indique que la finalité doit être "déterminée, légitime et explicite" et qu'elle "permet de déterminer la pertinence des données personnelles que vous recueillez". C'est donc une demande de la législation d'établir cette liste de finalité.

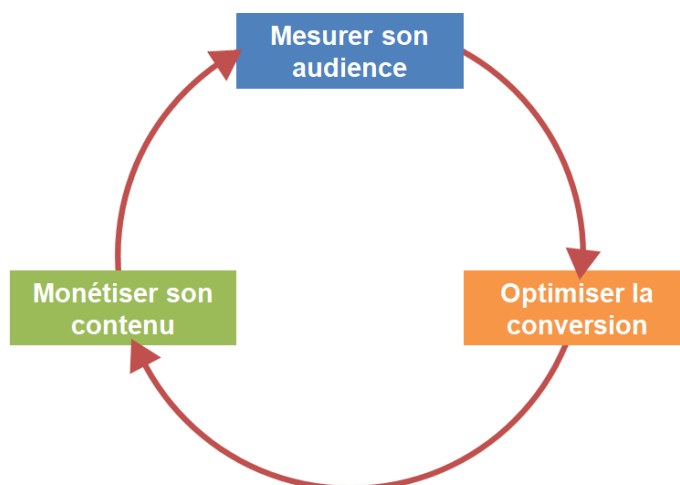


Schéma itératif de la collecte de données

1.1.1 Mesurer son audience

Les outils de mesure d'audience sont nombreux aujourd'hui sur le marché (Google Analytics, AT Internet, Adobe Analytics, Content Square, Session Cam, etc.) et lorsqu'un utilisateur navigue sur le Web, de multitudes d'informations circulent entre lui et le site internet. Nous pouvons les regrouper sous 3 grandes catégories : les données relatives aux utilisateurs, aux sources de trafic et aux comportements des internautes. En rassemblant toutes ces informations dans les solutions de mesure d'audience et grâce à l'analyse de statistiques obtenues, la connaissance clients et prospects augmente considérablement.

Si le site internet propose un module de création de compte via un email ou un compte social comme Facebook, l'utilisateur peut fournir directement des données personnelles comme le nom, le prénom, l'adresse email, etc. Ces informations sont enregistrées dans des outils de gestion de la relation client (en anglais Customer Relationship Management ou CRM). La majorité des outils d'analyse n'autorise pas les informations personnelles identifiables en clair (en anglais Personally Identifiable Information ou PII). Cependant, il est possible de les recouper en utilisant des clés de réconciliation (encodage de l'email par exemple). L'utilisateur peut apporter d'autres informations grâce à sa machine et son navigateur. Il peut s'agir de données démographiques (âge et sexe) grâce aux cookies tiers, géographiques (zone et langue) avec l'adresse IP ou technologiques avec le *fingerprinting* (navigateur, système d'exploitation, catégorie d'appareil). Connaître ces informations permet d'optimiser la performance du site pour être certain qu'il fonctionne avec la configuration de l'utilisateur.

Les sources de trafic désignent le moyen employé par un utilisateur pour venir sur un site internet. Grâce à des paramètres de campagnes dans l'URL de la page d'arrivée et/ou de la page qui précède l'arrivée sur le site (appelé également "referrer"), il est possible de définir, une fois la donnée collectée et traitée, des canaux d'attributions (Direct, Organique, Social, Email et Referral). Ces données sont particulièrement intéressantes car elles donnent une vision de la notoriété de la marque, portée par le site internet, sur le Web. Elles permettent également de prendre des décisions d'investissement sur des plateformes publicitaires comme le SEA (Search Engine Advertising) ou le social.

Les données comportementales mesurent l'engagement généré par un utilisateur envers un site Web. Elles sont relatives, par exemple, aux pages visitées et à leur temps de chargement, à la recherche de mots clés et aux interactions (clics, scroll, complétion de formulaire, etc.). Grâce à ces données, il est possible d'identifier le contenu qui plait le plus aux utilisateurs mais aussi de monitorer la performance du site (nombre de pages 404 rencontrées, boutons non fonctionnels, etc.). Pour un site e-commerce, les informations relatives aux intentions d'achat et aux transactions peuvent également être incluses dans cette catégorie comme par exemple la vue d'une liste de produit ou de promotions, l'affichage d'une page produit, l'ajout au panier, la navigation dans le tunnel de commande et la vue de la page de confirmation.

1.1.2 Optimiser la conversion

Une fois l'audience du site connue, il devient nécessaire d'optimiser les indicateurs de performance (taux de rebond, parcours, formulaire, etc.) pour proposer un contenu plus adapté aux visiteurs et optimiser le taux de conversion. Pour un site e-commerce, il peut s'agir d'augmenter son nombre de transactions et son chiffre d'affaires. Les tests A/B et la personnalisation contribuent à cette optimisation et dans ce domaine, les outils sont nombreux (Google Optimize, Kameleoon, AB Tasty, Optimizely, etc.).

Les A/B tests et les MV tests (Multiples Variantes) vont principalement utiliser les données comportementales acquises. Les premiers outils proposaient de comparer des versions de pages avec un seul élément modifié. Très vite, ils ont évolué et ont permis de modifier plusieurs éléments à la fois. En testant une ou plusieurs variantes d'une page (changement de couleur d'un bouton, d'un titre, de la position des blocs, etc.) auprès d'un segment d'audience sur une période définie et en collectant les résultats de l'expérience, il est possible de connaître la configuration du site la plus performante. La mise en place de ce type d'outil présente de nombreux avantages. Les équipes métiers peuvent, notamment, gérer directement leurs tests dans l'outil sans demande de modifications du code par les développeurs. Une fois qu'une variation est déterminée comme gagnante, elle peut être intégrée sur le site pour l'ensemble des utilisateurs.

La personnalisation exploite les données relatives aux utilisateurs. En connaissant leurs préférences d'achats, leur budget pour les transactions, leur tranche d'âge, leur genre, etc, il est possible de proposer un contenu adapté mis en avant sur le site (par exemple, via une bannière sur la page d'accueil affichant un produit en relation avec ce que l'utilisateur achète fréquemment). L'intérêt est d'accompagner rapidement l'internaute vers l'objectif de conversion. Moins il passe de temps à rechercher ce pourquoi il est venu sur le site, plus vite il sera satisfait, plus l'image de cette marque lui procurera un sentiment positif et plus il sera susceptible de revenir une prochaine fois.

1.1.3 Monétiser son contenu

En connaissant bien ses utilisateurs, il est possible de leur proposer de la publicité personnalisée diffusée sur plusieurs sites internet qu'ils consultent, grâce aux données des sources de trafic. Ces canaux d'attribution sont riches d'informations. Ils permettent de mesurer la notoriété de la marque et de l'accroître via un investissement de temps et parfois d'argent. Les solutions d'annonces publicitaires sont abondantes et comptent parmi elles Google Campaign Manager, Google Ads, Facebook Business Manager mais aussi Amazon Sizmek par exemple.

Un pourcentage important de trafic direct, associé aux utilisateurs accédant à un site internet en inscrivant l'URL directement dans la barre d'adresse ou via un onglet de la barre de favoris, est un bon indicateur de notoriété. Le site internet associé est reconnu. Néanmoins, la publicité

permet de réengager des clients existants et de convertir de potentiels prospects en nouveaux clients. Elle est présente sous différents formats et est suivie, dans le cas des réseaux sociaux, par une communauté à travers des posts fréquents.

La présence d'une marque sur des sites partenaires, dans le cadre de l'affiliation, augmente la confiance des internautes. Cette solution payante utilise le partenaire comme prescripteur et cible ses clients comme prospects de la marque. Par exemple, un site e-commerce proposant des t-shirts affichant des encarts publicitaires pour un club de sport.

1.2 L'émergence du cookie

Que ce soit pour connaître son audience, optimiser la conversion ou monétiser son contenu, les échanges de données marketing reposent principalement sur des cookies pour suivre l'utilisateur et son parcours sur plusieurs sites sur des périodes plus ou moins longues. Afin d'en tirer le plus de profit possible, les acteurs de la publicité ont exploité une faille existante dans la mise en place des cookies pour construire leur modèle économique.

1.2.1 La genèse

En 1994, la division serveur de la société Netscape² faisait face à un problème. Elle souhaitait transmettre les informations de navigation d'un utilisateur, sur un site internet, de page en page. L'exploitation des paramètres d'URL était envisagée mais complexe, notamment dans le cadre des informations de panier pour un site e-commerce. Les développeurs de Netscape ont alors créé les cookies pour répondre à ce besoin.

La CNIL définit un cookie comme "un petit fichier stocké par un serveur dans le terminal (ordinateur, téléphone, etc.) d'un utilisateur et associé à un domaine web"³. Les cookies sont utilisés pour de multiples raisons. Ils servent notamment à enregistrer des informations relatives à un utilisateur et son parcours (identifiant de connexion, préférences de langue, consentement ou refus face à la demande de collecte de données, etc.). Prenons l'exemple d'un utilisateur qui se rend sur un site e-commerce une première fois sans acheter et qui revient le lendemain pour finaliser sa commande.

² Fondée en 1994 et spécialisée dans l'informatique, Netscape est surtout connue pour son navigateur web nommé Netscape Navigator utilisé par la majorité des utilisateurs en 1996.

³ Définition du terme "cookie" : <https://www.cnil.fr/fr/definition/cookie>

JOUR 1



Un utilisateur crée un compte. Le serveur dépose un cookie utilisateur « 123 » sur le navigateur.



L'utilisateur ajoute un produit au panier. Le serveur dépose un cookie panier « 456 » sur le navigateur.

JOUR 2



L'utilisateur revient. Il se connecte au site et récupère son panier grâce aux cookies.

Dépôt de cookies pour un visiteur sur un site e-commerce

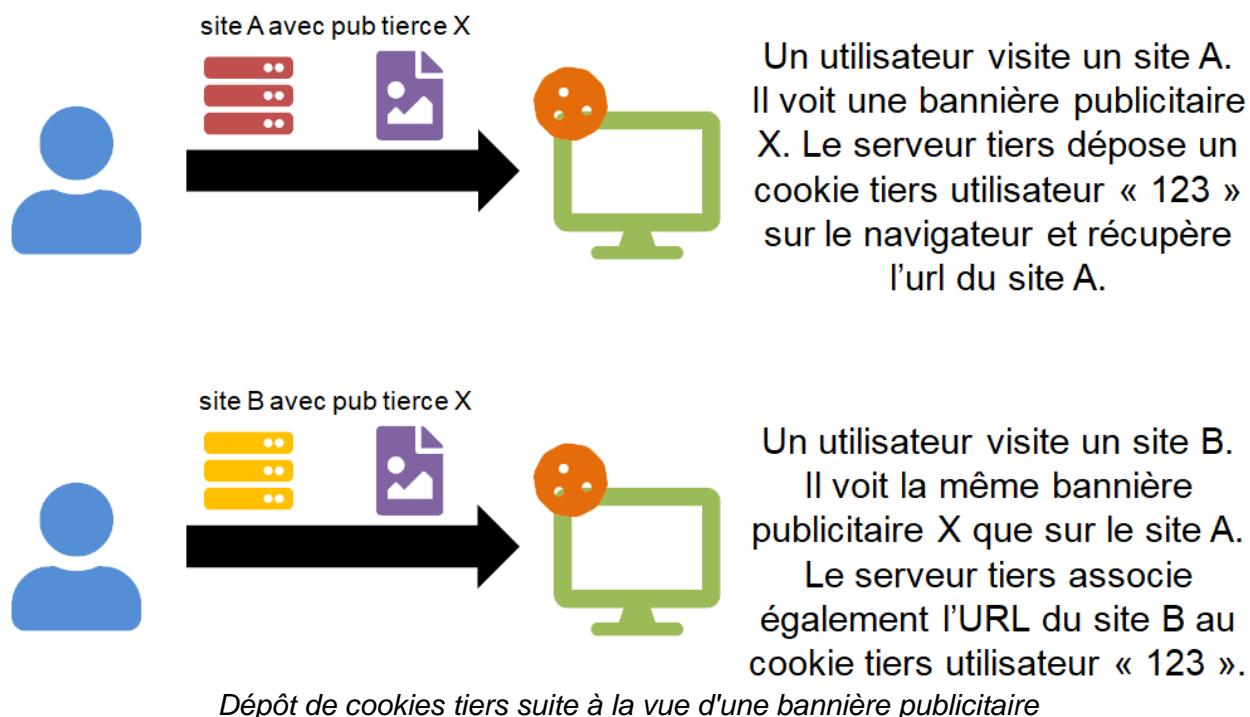
Le format des cookies fut développé précipitamment par Netscape. Le niveau de sécurité était faible et les cookies accessibles par d'autres serveurs. Un choix qui aura de lourdes conséquences sur la vie privée des utilisateurs.

1.2.2 La donnée personnelle en péril

Un site internet a la possibilité d'intégrer, dans son code, des images, des feuilles de styles, des iframes (des pages HTML imbriquées dans la page courante) ou des scripts par exemple. Ces ressources peuvent lui appartenir ou être appelées depuis un autre domaine. Et c'est là que le problème lié aux cookies fut pointé du doigt. Un site internet qui pose des cookies first (associés aux noms de domaines de la page en cours en opposition aux cookies third ou tiers associés à un domaine différent) et qui appelle des ressources tierces offre la possibilité aux serveurs tiers de lire et d'écrire des cookies first et tierce. Par conséquent, les données personnelles d'un utilisateur stockées sur son appareil via un cookie et liées à un site peuvent être partagées à d'autres sites qu'il visite si ces derniers appellent les mêmes ressources.

La fiche d'identité complète d'un utilisateur peut alors être établie, sans même qu'il le sache. Si un individu navigue sur un site A de vente de chaussures, un site B d'événements de courses à

pied et un site C de vente de vêtements de sport, et que ces 3 sites appellent la même ressource, le serveur tiers peut déterminer qu'il s'agit d'un sportif pratiquant la course à pied par exemple. C'est le principe de fonctionnement des bannières publicitaires. Ci-dessous, l'exemple d'un utilisateur qui visite un site A et un site B qui appellent tous les 2 la même bannière publicitaire tierce X.



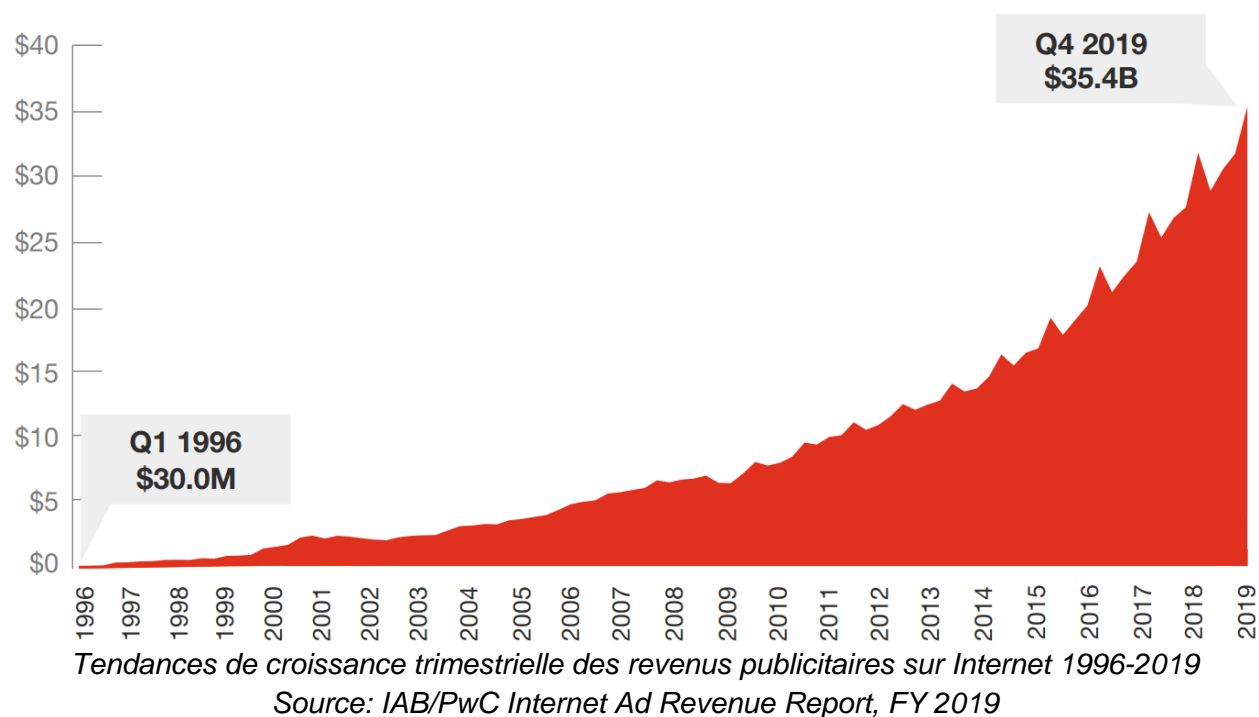
Lorsque les cookies se sont généralisés, le grand public n'a pas été informé de l'implémentation des cookies, de leurs contenus et du risque d'atteinte à leur vie privée. La première bannière publicitaire est apparue sur le site hotwired.com en 1994, soit la même année que la création des cookies. Elle a été conçue pour la société AT&T, spécialisée dans la téléphonie mobile américaine. Cette bannière est restée 3 mois en ligne sur le site et a généré un taux de clic de 44%. Ce qui fait donc du magazine américain Wired le précurseur de la publicité en ligne telle que nous la connaissons aujourd'hui.

1.2.3 La croissance d'un modèle économique

L'année suivante, c'est le moteur de recherche Yahoo qui intègre sur ses pages des espaces dédiés à la publicité. Le format des bannières instauré par hotwired.com était devenu une norme et fut repris par beaucoup d'acteurs comme IAB (Interactive Advertising Bureau) et W3C (World Wide Web Consortium). En 1996, les bannières sont devenues interactives. Hewlett Packard proposait de jouer à Pong d'Atari directement sur sa publicité. En 1999, l'affiliation fait son apparition. Les sites Web ont alors intégré les bannières promotionnelles de leurs partenaires sur leurs propres pages. Les vidéos publicitaires, avec Quicktime et Windows Media

Player, sont apparues cette même année. En 2000, Google lance les campagnes SEA avec AdWords.

Depuis les années 2000, l'économie liée aux publicités sur Internet s'est développée de manière exponentielle et offre de nombreux avantages que la publicité traditionnelle (TV, radio, presse) ne possède pas. Le format évolue (pop-up ou fenêtre secondaire, vidéo Flash, vidéo Youtube, annonce Facebook, mini-sites), le contenu est mieux ciblé, le coût est moins élevé et le retour sur investissement (ROI) plus important. Les annonceurs souhaitent intégrer de plus en plus la publicité avec les contenus éditorial et graphique de leurs sites. Un modèle économique entier repose sur les cookies et les acteurs l'ont très bien compris. À tel point qu'en 2017, les investisseurs placent davantage leur argent dans le digital que dans la télévision.



Aujourd'hui, la publicité est omniprésente dans notre navigation sur le Web. Sur Youtube, les vidéos que nous consultons sont coupées par de la publicité. Certains Youtubeurs font également de la promotion de produits directement dans leurs vidéos, ce qui "force" l'utilisateur à la voir. Nos données personnelles sont stockées dans des cookies, partagées entre plusieurs sites et parfois échangées contre de l'argent à certaines entreprises sans même que nous le sachions ou que nous ayons consenti à cette collecte de données. Ces données servent ou pourraient servir également à personnaliser encore plus le contenu publicitaire. C'est pourquoi le RGPD a décidé de siffler la fin de la récréation en publiant ses textes de lois.

1.3 La fin des cookies et ses conséquences

Actuellement, lorsque les entreprises souhaitent collecter de la donnée sur leur site web, elles sont confrontées à 3 grands mouvements.

Le premier est législatif et couvre les restrictions réglementaires imposées par le RGPD et la directive ePrivacy. La mise en conformité face à la loi devient la priorité des entreprises.

Le second est lié aux navigateurs qui peuvent empêcher un suivi et bloquer des comportements liés aux cookies, comme ITP (Intelligent Tracking Prevention) sur Safari, ETP (Enhanced Tracking Protection) sur Mozilla Firefox, le projet *Privacy Sandbox* de Google et le navigateur Brave qui bloque par défaut les traceurs.

Le troisième est lié aux extensions qu'on appelle "Ad blocker", comme Adblock, qui empêche des bannières publicitaires de s'afficher et de poser des cookies. Ces 3 mouvements mènent la vie dure aux cookies et aux traceurs et forcent les entreprises à repenser leur manière de collecter de la donnée.

1.3.1 Le RGPD et la directive ePrivacy dictent les règles du jeu

Le RGPD est le règlement de référence sur la protection des données à caractère personnel. Il est applicable dans les Etats membres de l'Union Européenne depuis le 23 mai 2018. Son objectif principal est le renforcement de la protection des données personnelles. Le RGPD offre un cadre juridique harmonisé pour les entreprises qui collectent et traitent ces données. Ce règlement responsabilise les entreprises et permet aux internautes de contrôler l'utilisation de leurs données.

La directive ePrivacy complète le RGPD et impose aux éditeurs de sites internet et d'applications mobiles de demander le consentement des utilisateurs avant de stocker ou d'accéder à des informations sur leurs appareils. Cette règle s'applique à tous les cookies et traceurs qui ne sont pas strictement nécessaires au bon fonctionnement du service proposé. L'article 4 paragraphe 11 du RGPD indique que le consentement doit être "libre, spécifique, éclairé et univoque". L'utilisateur doit pouvoir aussi facilement accepter que refuser la collecte de ses données.

Aujourd'hui, les sites internet et les CMP (Consent Management Platform) ne sont pas tous conformes au RGPD. Il n'y a pas d'harmonisation à l'échelle européenne et les diverses recommandations des organismes locaux peuvent donner lieu à plusieurs interprétations et niveaux de compréhension. Suite aux nouvelles recommandations de septembre 2020 proposées par la CNIL, en charge de la mise en application du RGPD et de la directive ePrivacy en France, les entreprises ont jusqu'au 31 mars 2021 pour se mettre en conformité sous peine de sanctions. Et le refus potentiel des utilisateurs face à la demande de consentement impacte grandement les entreprises et leurs partenaires. Un refus entraîne une perte de données sur les visiteurs et leur parcours pour les solutions analytiques et publicitaires, ce qui fausse le comptage réel dans l'interface. La réconciliation des parcours entre plusieurs appareils semble inconcevable si l'utilisateur refuse de partager ses données sur le site internet et l'application d'une même marque.

Même si la mise en conformité et la date limite de mars 2021 occupent les esprits des entreprises, les navigateurs qui intensifient leur chasse aux cookies ainsi que les extensions bloquant les publicités ont un impact non négligeable sur la collecte de données via les cookies.

1.3.2 Apple et la confidentialité des utilisateurs

Entre fin 2016 et 2017, plusieurs entreprises ont rencontré des failles de sécurité sur leurs applications et sites internet entraînant des fuites de données personnelles. Il y a eu notamment Instagram, avec 6 millions de comptes touchés, et Uber, avec 57 millions de comptes piratés. Mais la faille la plus importante revient à la société Équifax, spécialisée dans la cote de crédit à la consommation pour les entreprises et les particuliers, avec 143 millions de clients concernés. Suite à ces nombreux incidents et face à l'inquiétude et le manque de confiance grandissant des utilisateurs dans le traitement de leurs données personnelles sur Internet, Apple a décidé de réagir et de protéger les utilisateurs de Safari en déployant un programme nommé ITP (Intelligent Tracking Prevention).

La première version intitulée ITP 1.0 utilise le *machine learning* (l'apprentissage machine par l'intelligence artificielle) pour lister et catégoriser les domaines déposant des cookies afin d'identifier ceux qui le font pour un usage publicitaire et non pour un usage standard. ITP 1.0, face à ce problème, fait le choix d'instaurer une durée de vie maximale aux cookies tiers et de les supprimer automatiquement si l'utilisateur n'a pas interagi avec lui au-delà de 30 jours. Avec cette première version d'ITP, Apple déclare la guerre à ses concurrents (Google, Facebook, Criteo, etc.) et aux entreprises qui collectent de la donnée à des fins publicitaires.

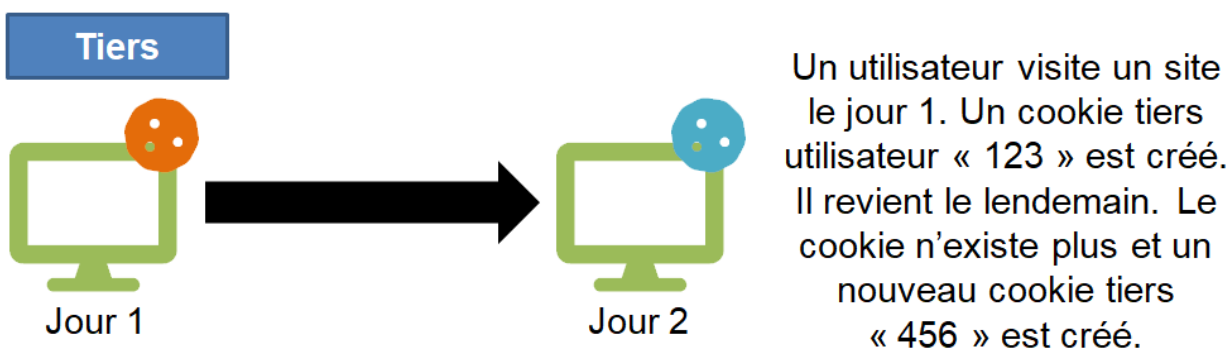
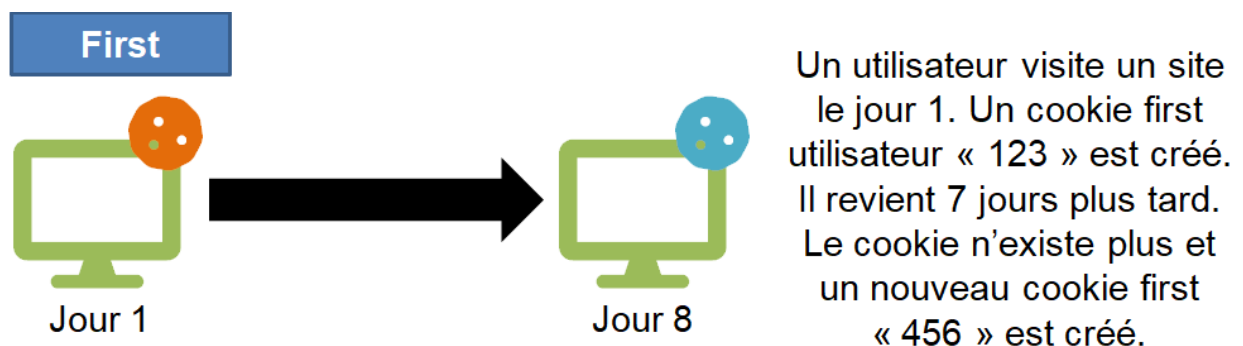
“Il a suffi qu'Apple restreigne la présence de cookies sur son navigateur Safari pour que Criteo plonge en Bourse en décembre.”

Source : Les Echos

Après avoir déployé plusieurs versions d'ITP et autant de méthodes de contournement imaginées par les développeurs, comme l'utilisation de cookies JavaScript et du *LocalStorage* pour stocker des informations sur le navigateur, Apple a sorti sa version 2.3 en septembre 2019 incluse à partir de Safari sur iOS 13 et Mac OS. Elle embarque les modifications des versions précédentes et en ajoute de nouvelles qui mènent la vie dure à la collecte de données. Le secteur du luxe, dont le trafic est important sur iOS, est l'un des plus touchés.

Dans cette dernière version drastique d'ITP, les cookies tiers ont une durée de vie de 24h, les cookies first créés en JavaScript expirent au bout de 7 jours, les paramètres de suivi dans les URLs (gclid, fbclid, etc.) sont limités à 24h et le LocalStorage est vidé au bout de 7 jours. Toutes ces contraintes ont un impact certain sur les données statistiques et publicitaires en empêchant de suivre correctement le parcours d'un utilisateur sur plusieurs jours. En effet, les solutions d'analyse, comme Google Analytics, déposent un cookie first tandis que les solutions médias, comme DoubleClick, déposent un cookie tiers pour attribuer un identifiant utilisateur. Il devient

impossible d'attribuer les conversions aux bonnes sources de trafic et de rémunérer correctement les affiliés.



Impact d'ITP sur les cookies first et les cookies tiers pour l'analytique et le média


Par conséquent, les annonceurs et les professionnels de la publicité doivent trouver des alternatives s'ils souhaitent continuer à collecter de la donnée sur Safari. Mais Apple n'est pas le seul à vouloir protéger ses utilisateurs. Google et Mozilla ont également leurs propres solutions pour leur navigateur.

1.3.3 Firefox et le consentement au niveau du navigateur

En octobre 2018, dans la version 63 du navigateur Firefox, Mozilla implémente la fonctionnalité nommée ETP (Enhanced Tracking Protection). Son objectif est de protéger la vie privée des utilisateurs durant leur navigation en bloquant une liste de traceurs et de cookies liés à des domaines connus pour le pistage, la publicité et les réseaux sociaux. Après plusieurs mises à jour, Firefox sort ETP 2.0 en juillet 2020 qui apporte un contrôle à 3 niveaux pour l'utilisateur (Standard, Stricte et Personnalisée) et des restrictions strictes pour les traceurs, comme le montre l'image ci-dessous.





Vie privée

Protection renforcée contre le pistage

 Les traqueurs vous suivent en ligne pour collecter des informations sur vos habitudes de navigation et vos centres d'intérêt. Firefox bloque un grand nombre de ces traqueurs et de scripts malveillants. [En savoir plus](#) [Gérer les exceptions...](#)

☐ **Standard** ▼
Équilibré entre protection et performances. Les pages se chargeront normalement.

☒ **Stricte**
Protection renforcée, mais certains sites ou contenus peuvent ne pas fonctionner correctement.

-  Traqueurs de réseaux sociaux
-  Cookies de pistage intersites
- ☒ Contenu utilisé pour le pistage dans toutes les fenêtres
-  Mineurs de cryptomonnaies
-  Détecteurs d'empreinte numérique

⚠ Attention !
Le blocage des traqueurs et l'isolation des cookies peut avoir une incidence sur les fonctionnalités de certains sites. Actualisez une page avec des traqueurs pour charger tout le contenu. [Me montrer comment faire](#)

☐ **Personnalisée** ▼
Choisissez les traqueurs et les scripts à bloquer.

Panneau de contrôle des différents niveaux de protection proposé par Firefox

Le niveau Standard d'ETP, qui est activé par défaut, bloque les traceurs et cookies connus par Firefox. Le site internet continue de fonctionner normalement. En revanche, avec le niveau Strict, la protection est renforcée et certains sites peuvent ne pas fonctionner correctement. Dans cette configuration, les requêtes à Google Analytics et à Facebook par exemple ne sont pas transmises à leurs serveurs. De plus, ETP 2.0 supprime au bout de 24h les données et les cookies de suivis des domaines connus. Cela empêche ces solutions de suivre les internautes de manière fine et d'établir un profil utilisateur sans connexion ou identifiant client.

1.3.4 Google et la promesse du respect de l'anonymat

Google intègre en novembre 2012, à partir de la version 23 de son navigateur Google Chrome, une fonctionnalité nommée Do Not Track (DNT). Cette dernière permet d'envoyer une demande d'interdiction de collecte de données et de suivi de l'utilisateur. Cette option n'est pas activée par défaut et est méconnue des internautes. De plus, les sites web n'ont pas d'obligations légales à prendre en compte cette requête. Certains sites internet cependant prennent en compte l'activation du DNT et respectent le choix de l'utilisateur. Dans ce cas, les solutions tierces ne se déclenchent pas et les cookies ne sont pas déposés. Néanmoins, l'utilisateur n'a aucune garantie que ses données ne soient collectées et utilisées sur plusieurs sites web.

www.urssaf.fr indique

You've enabled DNT, we're respecting your choice

OK

Affichage sur le site www.urssaf.fr d'une pop-up de respect de l'option DNT activée sur le navigateur Google Chrome

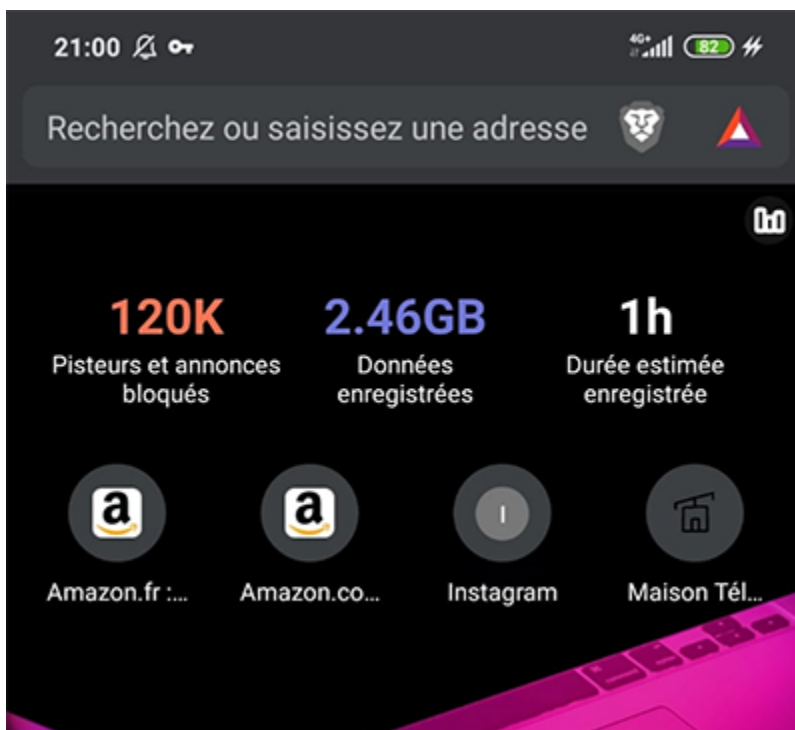
Face à la sous-exploitation du Do Not Track et aux fonctionnalités déployées par ses concurrents Apple et Mozilla pour renforcer la protection de données de leurs utilisateurs, Google a réagi et a annoncé en janvier 2020 son projet intitulé *Privacy Sandbox*. L'objectif fixé par Google d'ici deux ans est de continuer à mesurer la performance des annonces publicitaires tout en étant moins intrusif pour l'utilisateur.

Grâce à un système d'API, les cookies tiers publicitaires seront amenés à disparaître, ce qui va dans le sens des recommandations faites par la CNIL. De plus, le code sera proposé en open source. Il pourrait donc être adopté par d'autres navigateurs comme Firefox et Safari afin d'uniformiser les méthodes de collecte de données et éviter aux acteurs de développer des solutions différentes pour chaque navigateur.

En ce qui concerne le respect de l'anonymat de l'utilisateur, les informations de parcours seront stockées au niveau du navigateur. Les domaines tiers auront accès à des données restreintes et sélectionnées par la *Privacy Sandbox* de Google. Ce système de "boîte noire" obligera également les solutions tierces à mettre à jour leur code pour pouvoir communiquer avec les APIs proposées.

1.3.5 Brave et la protection de la vie privée

Brendan Eich, créateur du JavaScript et cofondateur de Mozilla, lance le navigateur web open source Brave en 2016. Son objectif avec Brave est la sécurité, la vitesse et le respect de la vie privée. Le navigateur bloque les traceurs, les cookies tiers et les publicités qui sont détectés comme solutions de collecte de données. Ce qui permet à l'utilisateur de protéger sa vie privée mais aussi d'améliorer la performance de sa navigation, la vitesse de chargement des pages et de prolonger la durée de vie des batteries de ses appareils (ordinateur portable, téléphone, etc.).



Affichage du nombre de publicités et d'annonces bloquées sur le navigateur Brave sur un smartphone après 2 ans d'activité

Brave se démarque surtout grâce à son modèle économique. Par défaut, le navigateur masque les publicités ciblées. A la place, il propose aléatoirement des notifications pour demander à l'utilisateur s'il accepte de visionner des annonces publicitaires non ciblées proposées par des partenaires de Brave. S'il accepte de regarder, le navigateur le récompense par de la cryptomonnaie intitulée BAT (Basic Attention Token) et qui se découpe en jetons. L'utilisateur peut ensuite faire des dons aux éditeurs de contenus qu'il apprécie le plus. Ce système peut être un vrai avantage pour des sites comme Wikipédia qui font très souvent des appels aux dons. En revanche, les cookies tiers et la publicité sont très impactés.

1.3.6 Les extensions bloquent la publicité

Les adblockers sont des logiciels et des extensions permettant à un utilisateur de bloquer l'affichage des publicités sur les sites internet lors de sa navigation. En inspectant le code source de la page, ils sont capables de détecter la présence de scripts appartenant à des annonceurs publicitaires connus et de les bloquer avant le chargement. Pour un utilisateur, il y a un vrai bénéfice à naviguer sans publicité : la vitesse de chargement des pages augmente, la bande passante est économisée et la visualisation du site est plus confortable.

En revanche, l'impact pour les solutions publicitaires semble évident. Si le contenu n'est pas chargé, les données liées à l'impression et au clic ne seront pas collectées et la régie publicitaire pourrait ne pas être rémunérée. Pour contrer ce problème, certains sites affichent une popin demandant la désactivation de l'adblocker pour accéder gratuitement au contenu et afficher la publicité.



Affichage d'une popin pour demander la désactivation des adblockers sur le site www.lesechos.fr

Mais l'un des adblockers qui affecte le plus le monde du cookie est uBlock Origin. En effet, ce dernier ne se contente pas de bloquer uniquement les publicités, il empêche une multitude de scripts de s'exécuter correctement. La solution de Google Analytics pour les statistiques, utilisant le domaine www.google-analytics.com, ne peut opérer correctement. Et de manière plus globale, le TMS⁴ Google Tag Manager, associé au domaine www.googletagmanager.com, est aussi bloqué par la solution. Ce qui implique que tous les scripts et les tags inclus dans le conteneur du TMS ne sont pas déclenchés.

⁴ Tag Management Systems (TMS en anglais) ou système de gestion des balises est un outil permettant à des personnes non techniques d'implémenter diverses solutions (analytiques, médias, a/b tests, etc.) sans intervention d'une équipe de développeurs.

1.4 Un enjeu de taille pour les entreprises

En plus des problèmes d'acquisition et d'exploitation des données liés aux limitations rencontrées par les différents cookies, il y a une réelle conséquence pour les entreprises dont la stratégie est conduite par la donnée. Pour connaître les impacts sur le quotidien de ces sociétés, j'ai créé un questionnaire composé de vingt questions. Quatre Web analysts, responsables des données collectées sur les sites internet et applications mobiles de leur entreprise, ont accepté de me répondre et de partager leur expérience face au RGPD et aux directives de la CNIL. Les questionnaires complets et anonymisés sont disponibles en annexe de ce mémoire. Trois principaux enjeux ont pu être identifiés grâce aux réponses : les pertes d'informations clients à limiter, la réorganisation des équipes pour suivre les évolutions législatives constantes et le coût de mise en place technique engagé.

1.4.1 Une connaissance client diminuée

Un des premiers impacts ressentis par les entreprises est une perte de données à collecter dû au refus des utilisateurs face au bandeau de demande de consentement aux cookies. D'une part sur la partie analytique, où il n'est pas possible de suivre le comportement du visiteur sur le site, et d'autre part sur le média. Un internaute venant d'une campagne média et refusant la collecte ne se verra pas attribuer le bon canal de conversion. Il devient donc difficile pour les entreprises de savoir exactement quel canal d'acquisition performe le mieux et d'ajuster leurs investissements médias en conséquence. La création et la réelle signification des audiences dans les interfaces de pilotage média sont donc moins précises. Les entreprises doivent donc mettre à jour leurs outils de collecte. Malheureusement, très peu de solutions sont aujourd'hui en mesure de fournir un outil répondant à toutes les attentes.

Les différents acteurs (analytique, média, personnalisation, etc.) n'ont pas tous réussi à anticiper les évolutions des recommandations de la CNIL de septembre 2020. Sur la partie analytique, quelques solutions ont su réagir en proposant une collecte de données anonymes. AT Internet avait déjà présenté avant cette date sa mesure hybride exemptée de consentement et validée par la CNIL. Côté Google, une fonctionnalité en bêta nommée Consent mode est sortie et permet d'ajuster le comportement des tags Google Analytics, Ads et Floodlight en fonction du consentement de l'utilisateur. S'il refuse une des finalités, Google va envoyer la donnée de manière anonyme aux solutions, en respectant les directives RGPD.

À l'inverse, les éditeurs médias n'ont pas devancé les nouvelles recommandations, ni opéré les changements nécessaires pour mettre en conformité leurs outils. Dans ce contexte évolutif, les solutions sont déjà très occupées à comprendre les tenants et aboutissants des mises à jour fréquentes des navigateurs comme Safari avec ITP ou ce qu'envisage Google avec *Privacy Sandbox*. Les solutions médias doivent trouver une méthode d'implémentation qui coche un maximum de cases législatives et techniques pour collecter les données en conformité.

1.4.2 Une organisation juridique

Afin de suivre la mise en conformité du RGPD, les entreprises ont dû mettre en place une structure juridique dédiée. Elle est principalement composée d'un Data Protection Officer (DPO) mais différentes personnes des départements marketing et digital peuvent compléter les effectifs. Si ces profils n'existent pas dans l'entreprise, il peut y avoir un potentiel coût de recrutement en plus du temps d'investissement pour lire et comprendre les textes de lois. Ce groupe juridique peut se réunir à une fréquence régulière pour suivre et garantir la bonne application des principales directives auprès des équipes concernées.

En interne, la structure juridique exerce une veille des recommandations émises par la CNIL et les autorités de protection des données (en anglais Data Protection Authority ou DPA) des autres pays sur le RGPD. Les webinars et diverses communications diffusées par les différents acteurs du marché complètent la connaissance des entreprises sur le sujet. Le DPO et les experts digitaux centralisent les informations relatives aux bonnes pratiques et créent des lignes directrices qu'ils diffusent auprès des équipes concernées. De plus, il y a un vrai travail d'annonce et de formation auprès des différents profils analytique et média pour leur expliquer la baisse de trafic, de session et d'utilisateurs dans les outils. L'analyse se fera uniquement sur un volume de visite pour lesquels les internautes ont donné leur consentement. Sauf si le DPO et son équipe décident de mettre en place une mesure anonyme exemptée et dans ce cas, il faudra expliquer que la réconciliation de données ne sera plus possible.

1.4.3 Un investissement technique

Bien que la CNIL ait mis à jour ses recommandations en septembre 2020, les entreprises ont anticipé la mise en place de ces nouvelles règles pour la collecte du consentement sur les cookies. La plupart d'entre elles possédait donc déjà un bandeau ou une popin de recueil du consentement. Néanmoins, les nouvelles recommandations changent la donne et impliquent de modifier sa CMP.

La mise en place d'une bannière de consentement a un coût de développement à prendre en compte et doit être suivi via un planning. Les entreprises sont amenées à mettre en place des A/B tests sur le format et la position du bandeau et des boutons, le contenu ou encore les couleurs. A la fin des tests, il est possible de connaître la version qui a la meilleure performance avec le taux d'acceptation le plus important. Un coût d'accompagnement par des cabinets de conseil doit être considéré pour aider à implémenter la CMP et établir les A/B tests. Les raisons sont nombreuses pour les entreprises mais presque obligatoires pour être conformes au RGPD.

Des outils variés vont continuer à évoluer et d'autres vont émerger suite aux implémentations en cours sur les sites et applications des entreprises. Un coût supplémentaire de développement pourrait donc s'ajouter aux coûts déjà évoqués. Les entreprises s'interrogent également sur des solutions pour tenter de récupérer les données non consenties via des logs ou d'estimer celles qu'ils auraient dû avoir via la modélisation. D'autres solutions techniques sont également envisagées pour récupérer un niveau de données semblable au RGPD.

Depuis 1994 et l'apparition du premier cookie, l'économie digitale a beaucoup évolué. Les cookies sont au centre de beaucoup de solutions, dont les publicités. Suite à une collecte intense des données personnelles des internautes par les solutions, le RGPD et les navigateurs s'engagent à renforcer la protection des utilisateurs. Les entreprises et les éditeurs doivent donc composer avec les obligations législatives et techniques tout en parvenant à collecter des données pertinentes pour leur secteur.

Dans la seconde partie, nous allons étudier les avantages et inconvénients de quatre solutions différentes : l'exemption CNIL, le server-side, la navigation authentifiée et la modélisation pour le *machine learning*. Nous tenterons d'identifier comment elles peuvent permettre une conformité au RGPD tout en conservant une qualité de service équivalente à la période précédant sa mise en place.

Partie 2 - Concilier respect du RGPD et exhaustivité de la connaissance utilisateur

Le monde sans cookies qui se prépare impacte aussi bien les acteurs de solutions que les entreprises qui collectent la donnée. La position de l'ICO (Information Commissioner's Office), qui est le pendant britannique de la CNIL, fait polémique car elle requiert un consentement explicite pour le dépôt de cookie à des fins purement statistiques⁵ tandis que la CNIL propose une exemption de consentement pour cette finalité sous certaines conditions. La CNIL dans ses recommandations demande également un consentement pour les tests A/B et la personnalisation. Et, dans le cas où l'utilisateur refuse tous les cookies, il devient non mesuré, c'est-à-dire parfaitement inexistant dans les outils de suivi.

Alors la question se pose : quelle solution permettrait de concilier qualité de service et respect du RGPD ? Les textes du RGPD sont souvent mis à jour et soumis à interprétation de la part des éditeurs de sites européens : durée de vie des cookies selon la réponse de l'utilisateur face à la bannière de consentement (l'acceptation pourrait être conservée plus longtemps et par conséquent le bandeau réaffiché plus rapidement dans le cadre d'un refus). Sans mesure précise de l'efficacité des parcours, des fonctionnalités, sans tests A/B ni personnalisation, la qualité de service des sites Web européens ne peut rester au niveau. Les éditeurs de sites non européens risquent alors d'être avantagés car ils pourront apprendre de leurs données provenant d'utilisateurs en dehors de l'Europe sans encombre afin de construire et d'optimiser les interfaces de manière efficace.

⁵ Décision de l'ICO sur l'exemption des cookies analytiques : <https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/how-do-we-comply-with-the-cookie-rules/#comply15>

Les différents acteurs et entreprises vont donc être contraints de réagir et d'engager une réflexion sur les moyens possibles pour continuer à perfectionner leurs connaissances utilisateurs. Dans cette partie, nous allons étudier les avantages et inconvénients de solutions disponibles pour obtenir un maximum de données tout en respectant la conformité au RGPD.

2.1 Exemption CNIL

Le RGPD et la directive ePrivacy imposent de recueillir le consentement de l'utilisateur avant de pouvoir déposer des cookies sur son terminal. Néanmoins, la CNIL indique dans ses recommandations⁶ que certains traceurs de mesure d'audience, utilisés pour obtenir des informations sur la performance, la fréquentation et les contenus consultés d'un site internet, peuvent bénéficier d'une exemption de consentement sous certaines conditions. Les recommandations de la CNIL s'appliquent uniquement pour la France. Un éditeur de site multi pays doit donc se renseigner auprès des organismes dédiés de chaque pays dans lequel il est présent pour vérifier s'il peut également exempter la mesure d'audience dans le pays en question.

2.1.1 Collecte de données dès la première page

Grâce à cette exemption, les outils de mesure d'audience qui en bénéficient peuvent déposer un cookie dès la première page vue lors de la navigation de l'internaute. Cette première page vue est cruciale car elle permet de récupérer les taux de rebonds et les sources de trafic pour mesurer les performances des campagnes. Cette collecte de données n'étant pas désactivable par l'utilisateur, elle donne l'avantage de pouvoir suivre l'ensemble de son parcours sur le site internet.

Le bandeau de consentement peut alors être facultatif si seule la solution de mesure exemptée est implémentée. L'éditeur du site devra néanmoins informer les internautes de la mise en place de ce traceur via sa politique de confidentialité. La transparence et le respect de l'anonymat permettent de renforcer l'identité de la marque et la confiance de l'utilisateur.

2.1.2 Une finalité limitée

La CNIL est claire sur la façon dont l'outil implémenté doit être utilisé pour bénéficier de l'exemption en disant qu'il doit "avoir une finalité strictement limitée à la seule mesure de l'audience du site"⁵. Par conséquent, toute autre solution ne vérifiant pas cette règle ne peut bénéficier de l'exemption CNIL et doit être soumise à l'acceptation du consentement par l'internaute pour transmettre de l'information. Il est impossible, par exemple, d'utiliser la donnée collectée dans le cadre de la mesure d'audience pour définir des groupes d'utilisateurs selon des critères spécifiques (segmentation), du retargeting ou de la personnalisation. Elle doit servir

⁶ Bénéficier de l'exemption de consentement selon la CNIL : <https://www.cnil.fr/fr/mesurer-la-frequentation-de-vos-sites-web-et-de-vos-applications>

uniquement à des fins statistiques anonymes et rester hermétique à tout recoupement avec d'autres sources de données. La connaissance client par l'entreprise devient donc moins précise.

Une autre condition à respecter pour bénéficier de l'exemption proposée par la CNIL est que l'outil d'audience ne doit pas "permettre le suivi global de la navigation de la personne utilisant différentes applications ou naviguant sur différents sites web"⁷. Ainsi, il ne sera plus possible de faire des analyses entre différents domaines. Si un éditeur de site propose un parcours interdomaine à ses internautes, par exemple www.monsite.fr vers www.monsite.es dans le cadre d'un changement de langue ou www.monsite.fr vers moncompte.monsite.fr pour afficher le profil, il ne pourra pas réconcilier les données des différents domaines entre elles. Les développeurs auront un coût potentiel et devront repenser l'architecture du site s'ils souhaitent obtenir l'exemption CNIL et suivre ce type de parcours (www.monsite.com/fr vers www.monsite.com/es ou www.monsite.com/mon-compte).

2.2 Server-side

En règle générale, les implémentations de scripts pour collecter et transmettre de la donnée vers des outils tiers sont faites côté client (client-side). Elles peuvent être embarquées au sein d'un TMS ou être implémentées directement dans le code source du site. Les requêtes générées vont être visibles sur le navigateur de l'utilisateur via l'onglet Réseau (Network en anglais) de l'outil de développement. À l'inverse, l'implémentation d'une logique *server-side* va déplacer tout ce travail sur un serveur distant. Cette technique offre de nombreux avantages quant à la performance du site et permet d'échapper aux restrictions de suivi imposées par certains navigateurs. Néanmoins, cette méthode nécessite toujours une demande de consentement de l'utilisateur pour la collecte de données.

2.2.1 Réduire la charge client pour un meilleur contrôle

En déportant le chargement des scripts *server-side*, les pages contiennent moins de lignes de code à lire et moins de scripts à déclencher. Par conséquent, la performance du site augmente, les risques qu'une erreur dans le script fasse buguer ou crasher le site sont amoindris. Les navigateurs qui mettent en place un blocage des systèmes de suivis comme ITP et ETP, ainsi que les adblockers, ne sont pas en mesure d'empêcher l'envoi de requêtes du serveur vers les solutions tierces.

Un cas rencontré souvent par les équipes concernées lors des implémentations côté client, notamment sur un site e-commerce, est l'écart du nombre de transactions, et donc de revenus, entre l'outil d'analyse et le back-office. Ceci est dû au temps nécessaire au chargement du script de la solution d'analyse. Si l'utilisateur quitte le site avant que le script se charge, la donnée n'est pas envoyée. Le back-office est plus fiable et plus précis, ce qui oblige l'analyste à

⁷ Ne pas suivre la navigation d'une personne sur différentes applications ou sites web : <https://www.cnil.fr/fr/mesurer-la-frequentation-de-vos-sites-web-et-de-vos-applications>

jongler avec les deux outils. Grâce au tracking *server-side*, la solution analytique se déclenche côté serveur une fois la confirmation de paiement reçue et est indépendante de la fermeture de la page par l'utilisateur sur le site. Le nombre de conversions est donc égal à celui rapporté par le back-office.

Les données transmises par les développeurs au serveur et envoyées aux diverses solutions sont sous contrôle avec cette méthode. Les scripts appelés par *piggybacking* (appels de tags imbriqués au sein d'autres tags à l'insu du site qui les implémente et parfois envoyées sans consentement) et *data leakage* (fuite de données vers des tiers non désirés) ne peuvent pas récupérer les informations du site internet car exécutés côté serveur. Il y a également une homogénéisation des informations. Il est possible de récupérer l'adresse IP calculée par le serveur pour la transmettre à chaque solution.

2.2.2 Consentement et temps de développement

Un tracking *server-side* n'implique pas une exemption de la loi. Il est impératif d'obtenir le consentement de l'utilisateur avant de déclencher l'envoi de requête vers diverses solutions. Les développeurs du site internet doivent récupérer les informations de la CMP et les transmettre dans la requête au serveur. Il sera ainsi possible de conditionner les scripts en fonction du choix de l'internaute.

L'utilisation du *measurement protocol*, qui permet aux développeurs de faire des requêtes HTTP pour envoyer des données brutes d'interaction utilisateur directement aux serveurs de la solution, est la méthode la plus utilisée pour le suivi en *server-side*. Malheureusement, elle n'est pas supportée par toutes les solutions. Le choix d'une implémentation en *server-side* dépend donc des outils de destinations choisis.

Un effort supplémentaire est demandé aux développeurs avec le tracking en *server-side* car toutes les informations calculées et transmises naturellement par les bibliothèques JavaScript ne le sont plus. L'identifiant de session, le user-agent, la géolocalisation, les paramètres de campagne ou encore le *timestamp* pour analyser l'utilisateur et son parcours sous plusieurs points de vues sont toujours nécessaires. Le plan de mesure et de collecte de données à fournir aux développeurs doit être plus détaillé pour prendre ces éléments en considération.

2.3 Navigation authentifiée

De nombreux sites internet proposent à leurs visiteurs de créer un compte utilisateur. Que ce soit pour commander des produits sur un site marchand, s'abonner à un service de presse en ligne ou encore à un réseau social. Certaines plateformes nécessitent d'ailleurs d'être membre pour accéder au contenu qu'elles proposent, comme pour les acteurs de la vente privée (Voyage Privé, Veepee, Showroomprive, etc..). Il n'est pas possible de connaître leurs offres sans se connecter. Par la suite, les utilisateurs bénéficient des fonctionnalités liées à leur profil, leur panier et leurs commandes en échange de leurs informations personnelles. Par exemple,

une adresse email pour être prévenu de l'ouverture d'une vente chez Veepee. Le site internet peut alors personnaliser la relation avec son client. En revanche, il ne doit pas recouper ces informations avec d'autres solutions analytiques et médias sans le consentement de l'utilisateur.

2.3.1 Une relation personnalisée

En proposant une création de compte, le site internet récupère de nombreuses informations personnelles sur ses utilisateurs (nom, prénom, âge, email, etc...) et plus la demande de création de compte arrive tôt dans le parcours de l'utilisateur, plus rapidement le site s'assure de récupérer ces données. L'éditeur du site devient alors propriétaire de ces données. Il peut les exploiter selon son bon vouloir. Ainsi, par exemple, si le site internet propose un parcours e-commerce, les informations de commande pourront être utilisées pour cibler les utilisateurs sur le canal de contact où le client est le plus réactif et lui proposer des produits ou des promotions en relation avec son achat et ses préférences. Cela peut prendre plusieurs formes : emailings avec l'email de l'utilisateur, sms grâce à son numéro de téléphone portable, campagnes de mailing à l'adresse de livraison ou encore pushes notifications sur les téléphones dans la cas des applications. À ce titre, on conçoit aisément que la stabilité de l'identifiant pour suivre un utilisateur dans la relation est importante.

Lorsque les données CRM sont réconciliées avec des données de mesures d'audience grâce à l'identifiant de compte, la connaissance de l'utilisateur est fortement enrichie. Il est possible de comprendre dans les détails son parcours et de le personnaliser en fonction de son profil. Dans son outil analytique, Google propose une fonctionnalité nommée "Google Signals". Elle permet de croiser les données du site internet et de l'application mobile si l'utilisateur s'est connecté sur les deux appareils avec le même compte afin de lui proposer de la publicité personnalisée.

Si l'on prend en compte la complexité, somme toute relative, que nécessite la mémorisation d'un mot de passe supplémentaire côté utilisateur, le compte client apparaît alors comme un espace sécurisé pour centraliser son identité et ses données. Il peut y retrouver ses informations, ses historiques de commandes, les produits qu'il préfère mais également les avantages qu'il cumule avec une carte de fidélité. Sa navigation devient personnalisée. Elle reflète ses goûts et lui permet de gagner un temps précieux. Une relation privilégiée s'instaure donc entre l'utilisateur et la plateforme, comme lorsque l'on entre dans une boutique et que le vendeur nous appelle par notre prénom.

2.3.2 Frein potentiel à la conversion de l'utilisateur

Il n'y a pas vraiment d'inconvénient pour une entreprise à proposer une création de compte à ses visiteurs si elle est pertinente. La demande d'inscription doit être présentée au bon moment en échange d'une promesse de valeur ajoutée (une remise sur le prochain achat, retrouver ses commandes, etc.). Dans le cas contraire, il y a un risque potentiel de faire fuir l'internaute. D'ailleurs, certains sites comme les portfolios n'ont ni l'intérêt ni l'utilité à mettre en place un module de création de compte. L'entreprise a une part de responsabilité vis-à-vis des données collectées. En plus de ne pas collecter des données qui n'ont pas de cohérence avec la fonction

du site, elle doit également permettre à l'utilisateur de pouvoir supprimer toutes les informations obtenues sur lui de la manière la plus simple possible mais aussi être en mesure de prouver qu'il ne reste aucune trace de lui. En ce qui concerne la réconciliation des données CRM avec d'autres solutions, elle est possible uniquement si le consentement de l'utilisateur est obtenu.

L'expérience utilisateur avec la création d'un compte doit être maîtrisée par l'entreprise. Un formulaire qui demande beaucoup d'informations peut rendre réticents les prospects. D'ailleurs seules les informations strictement nécessaires peuvent être demandées dans un premier temps. Les sites e-commerces ont d'ailleurs vu apparaître des tunnels d'achats en mode invité, n'obligeant pas l'utilisateur à s'inscrire et lui permettant tout de même de passer commande. Forcer la création du compte entraîne des abandons de panier et une perte de prospects qui auraient pu ne plus avoir envie de visiter le site.

2.4 Modélisation et Machine Learning

Si un utilisateur refuse la collecte de données, via le bandeau ou la popin de demande de consentement, il ne peut pas être suivi. De ce fait, plusieurs indicateurs de performance, comme le nombre d'utilisateurs et de sessions, sont alors impactés car les informations ne sont pas disponibles dans les outils d'analyse.

Grâce au *machine learning* et aux données des utilisateurs qui ont consenti, il est possible toutefois de simuler le comportement des utilisateurs non mesurés. Et avec la modélisation qui en découle, établir la structure des informations obtenues pour estimer des indicateurs dans un scénario où tous les utilisateurs auraient accepté les cookies de mesure.

2.4.1 Construire des données

Cette méthode permet d'obtenir des estimations liées aux indicateurs de performance clés pour les équipes dédiées tout en étant conforme au RGPD. La machine apprend et modélise à partir des données de navigation fournies par les utilisateurs qui ont accepté les cookies de mesure. Elles peuvent ensuite être utilisées pour restituer, dans une interface graphique, les résultats. Les entreprises pilotées par la donnée peuvent ainsi continuer à prendre des décisions tout en respectant le choix de leurs utilisateurs.

Si le site internet possède des données collectées avant la mise en place de la demande de consentement, elles couvrent donc l'exhaustivité des utilisateurs. Elles peuvent alors servir de base de référence pour le machine learning et la modélisation et être comparées avec le résultat obtenu par l'estimation.

2.4.2 Estimation approximative

Cette solution reste toutefois une estimation approximative du nombre d'utilisateurs, de sessions et d'interactions. Même si certaines solutions de CMP permettent d'obtenir des statistiques sur l'utilisation de leur outil de consentement et notamment du nombre de refus, il n'y a aucune assurance concernant la fiabilité du volume d'utilisateurs n'ayant pas accepté. Il est nécessaire d'avoir un volume suffisant de données pour utiliser le machine learning. Les résultats risqueraient, dans le cas contraire, d'être complètement faussés.

Le temps de préparation, la capacité des machines et le retraitement des données sont des informations cruciales dans le coût de la mise en place. Chaque changement significatif du site internet, modifiant le comportement de l'utilisateur, implique un nouvel entraînement ou une nouvelle itération pour la machine. Il peut donc être complexe d'instaurer et de maintenir cette méthode.

2.5 Un consentement souvent nécessaire

Les solutions que nous venons d'aborder ont chacune leur lot d'avantages et d'inconvénients. Elles permettent soit de mesurer l'audience d'un site internet de manière anonyme et sans consentement soit d'obtenir des informations personnelles des utilisateurs après acceptation d'un bandeau ePrivacy. Aujourd'hui, aucune solution complète permettant une connaissance accrue de ses utilisateurs en toute conformité au RGPD existe. C'est aux entreprises de choisir celle qui correspond le mieux à leur situation en fonction de différents critères (faisabilité technique, coût d'intégration, compétences internes, etc).

Dans le cadre d'une mesure d'audience anonyme, il est possible d'obtenir une exemption de la CNIL. Mais si le site internet souhaite mesurer une audience identifiée et a implémenté d'autres outils, le consentement de l'utilisateur devient alors indispensable pour réconcilier les données. C'est pourquoi, les entreprises investissent du temps dans des A/B tests de bandeaux pour accroître leur chance d'obtenir une acceptation totale aux différentes finalités énoncées.

La réponse à la question de la réconciliation entre le respect du RGPD et le maintien de la qualité de service proposée n'est pas forcément l'application d'une seule et unique solution. Elle est peut-être la combinaison de solutions évoquées et d'autres solutions innovantes qui vont émerger par la suite. Toutes ces solutions n'excluent pas forcément la demande de consentement pour la collecte des données de l'utilisateur. Pour mieux comprendre les enjeux liés à l'implémentation de ces solutions, nous allons expérimenter dans la partie suivante une collecte de données server-side et conforme au RGPD.

Partie 3 - Expérimentation d'une collecte de données sans cookies, sans JavaScript et conforme au RGPD

Les cookies ayant la vie dure, les entreprises cherchent et testent des solutions de collecte de données leur permettant de concilier la conformité au RGPD avec une connaissance fine de leurs utilisateurs. Précédemment, nous avons évoqué plusieurs solutions en mettant en lumière leurs avantages et inconvénients respectifs. Nous allons donc tenter désormais l'expérience suivante : la mise en place d'un outil de mesure d'audience sans cookies et conforme au RGPD. Dans une volonté d'efficacité et afin de limiter au maximum l'impact sur les performances du navigateur des internautes, la solution que nous proposerons sera sans JavaScript et les requêtes seront traitées côté serveur.

Pour mettre en œuvre cette expérience, j'ai créé un site internet non commercial, mettant en page mon curriculum vitae, et accessible à l'adresse suivante : www.quillaume-sinnaeve.com. L'objectif de ce projet est de comprendre la méthode utilisée pour collecter de la donnée et d'analyser les informations récupérées sur les visiteurs. Dans cette partie, nous aborderons dans un premier temps les obligations législatives à respecter pour collecter la donnée des utilisateurs dans le cadre de l'exemption CNIL. Ensuite, nous verrons les solutions techniques utilisées pour suivre le parcours du visiteur sur le site sans cookies et sans JavaScript, grâce à d'autres langages et un paramètre d'URL. Puis, nous aborderons le traitement effectué sur les données pour faciliter leur lecture et leur stockage dans la base. Enfin, nous mettrons en place une page d'analyse contenant des visualisations graphiques des données collectées.

3.1 Les pré-requis législatifs

Dans le cadre de cette expérience, j'ai opté pour le choix d'une mesure d'audience exemptée de consentement selon les recommandations de la CNIL. Cela me permet de récupérer les informations des visites et d'éprouver l'outil de collecte de données mis en place. Le site étant une vitrine de mon parcours et n'ayant pas d'objectifs de vente, il est principalement destiné à des internautes français. Nous allons donc étudier et implémenter les recommandations émises par la CNIL pour être en conformité et bénéficier de l'exemption de consentement.

3.1.1 Une finalité stricte

L'exemption CNIL implique le respect par l'éditeur du site de quelques règles bien définies. Une des plus importantes est la finalité. Elle doit être "strictement limitée à la seule mesure de l'audience du site ou de l'application [...] pour le compte exclusif de l'éditeur"⁸. Pour notre expérience, l'outil mis en place a pour objectif d'analyser de manière anonyme les contenus consultés par les visiteurs. Cette solution ne posant pas de cookies et étant développée uniquement sur le site www.quillaume-sinnaeve.com, elle ne permet pas de suivre les visiteurs

⁸ La mesure d'audience pour le compte exclusif de l'éditeur : <https://www.cnil.fr/fr/mesurer-la-frequentation-de-vos-sites-web-et-de-vos-applications>

à travers différents sites internet. En ce qui concerne le non recoupement et le non transfert des données vers des solutions tierces, c'est à l'éditeur du site d'assurer l'application de cette règle. Les visiteurs n'ont aucune garantie et doivent faire confiance à l'éditeur. Dans le cadre de l'expérience, la solution n'est pas recoupée et les informations sont stockées sur une base de données propre au site (voir Annexe 8) et non accessible à des tiers.

#	Nom	Type	Interclassement	Attributs	Null	Valeur par défaut
1	timestamp	timestamp			Non	CURRENT_TIMESTAMP
2	session_id	varchar(21)	utf8_general_ci		Oui	NULL
3	event_name	varchar(40)	utf8_general_ci		Oui	NULL
4	event_custom_1	varchar(100)	utf8_general_ci		Oui	NULL
5	event_custom_2	varchar(100)	utf8_general_ci		Oui	NULL
6	page_uri	varchar(512)	utf8_general_ci		Oui	NULL
7	date_full	varchar(10)	utf8_general_ci		Oui	NULL
8	date_day	varchar(16)	utf8_general_ci		Oui	NULL
9	date_hour	int(2)			Oui	NULL
10	date_minute	int(2)			Oui	NULL
11	referrer_full	varchar(512)	utf8_general_ci		Oui	NULL
12	referrer_host	varchar(128)	utf8_general_ci		Oui	NULL
13	referrer_path	varchar(128)	utf8_general_ci		Oui	NULL
14	geo_country	varchar(128)	utf8_general_ci		Oui	NULL
15	geo_region	varchar(128)	utf8_general_ci		Oui	NULL
16	geo_city	varchar(128)	utf8_general_ci		Oui	NULL
17	device_userAgent	varchar(256)	utf8_general_ci		Oui	NULL
18	device_category	varchar(32)	utf8_general_ci		Oui	NULL
19	device_system	varchar(32)	utf8_general_ci		Oui	NULL
20	device_browser	varchar(32)	utf8_general_ci		Oui	NULL
21	device_language	varchar(32)	utf8_general_ci		Oui	NULL

Schéma de la base de données collectées dans le cadre de l'expérience

Les entreprises ont un devoir d'information envers les utilisateurs. C'est pourquoi, le site doit disposer d'un texte faisant référence à la finalité des données collectées. Cette mention peut être présente sur un bandeau de consentement mais aussi sur une page dédiée à la politique de confidentialité.

3.1.2 Informer les utilisateurs

Le RGPD repose sur la transparence des éditeurs de site internet avec leurs utilisateurs. Il encourage à expliciter au maximum les traitements faits sur les données à caractère personnel. Dans le cadre de l'exemption CNIL, il n'est pas nécessaire de mettre en place un bandeau de consentement s'il s'agit de la seule solution implémentée sur le site. Néanmoins, il est indispensable d'expliquer aux utilisateurs la finalité du traitement effectué. Une mention au sein d'une page de politique de confidentialité⁹ répond à ce besoin (voir Annexe 5). Cette page doit être accessible à tout moment sur le site. Par conséquent, j'ai opté pour la présence d'un lien de redirection vers la politique de confidentialité au bas de toutes les pages.

Données personnelles

Sur le site web www.guillaume-sinnaeve.com, il y a 1 type de données susceptible d'être recueilli :

- **Les données collectées automatiquement**
Lors de vos visites, je recueille des informations de type « web analytics » relatives à votre navigation comme la durée de votre consultation, votre localisation, votre type et version de navigateur. La technologie utilisée est sans cookies et sans JavaScript.

Mention de la finalité des données collectées au sein de la page de politique de confidentialité

Concernant la durée de vie des cookies et autres traceurs, la CNIL recommande un maximum de treize mois. Sur le site utilisé pour l'expérience, il n'y a pas de cookies. En substitution, c'est un paramètre d'URL "sid" qui effectue le suivi du visiteur sur différentes pages. La valeur de ce paramètre est renouvelée au-delà de 30 minutes d'inactivité du visiteur. Ci-après un exemple d'URL contenant le paramètre "sid" avec une des valeurs attendues :

https://www.guillaume-sinnaeve.com/fr/home?sid=1609693795_275788320.

Toutes les informations qui sont collectées dans le cadre de l'expérience sont stockées sur une base de données. La CNIL recommande une conservation maximale d'une durée de vingt-cinq mois. Pour être certain de ne pas détenir des données au-delà de cette limite, une routine a été mise en place sur la base de données (voir Annexe 6). Quotidiennement, un contrôle de la base de données est effectué pour supprimer les informations dont la date d'enregistrement est supérieure à 25 mois.

⁹ Pages sur la politique de confidentialité : <https://www.guillaume-sinnaeve.com/fr/privacy-policy> et <https://www.guillaume-sinnaeve.com/en/privacy-policy>

Nom de l'évènement	DELETE_DATA_OVER_25_MONTHS	
État	ENABLED ▼	
Type d'évènement	RECURRING ▼	
Exécuter chaque	1	DAY ▼
Départ	<input type="text"/>	
Fin	<input type="text"/>	
Définition	<pre>1 DELETE FROM `data` WHERE `timestamp` < CURRENT_TIMESTAMP - INTERVAL 25 MONTH</pre>	

Création de l'événement de suppression automatique des données enregistrées au delà de 25 mois dans la base

Nous connaissons désormais les pré-requis législatifs à respecter pour mettre en place une collecte de données sur le site www.guillaume-sinnaeve.com dans le cadre de l'expérience tout en étant conforme au RGPD. Nous allons maintenant nous intéresser aux contraintes techniques imposées pour cette expérimentation et aux méthodes utilisées pour réussir la collecte.

3.2 Les contraintes techniques

Dans le cadre de ce projet de mesure d'audience du site, plusieurs règles sont imposées : l'utilisation du JavaScript est proscrite, le traitement des informations doit être effectué côté serveur et aucun cookie ne doit être créé.

L'objectif étant de capitaliser sur les langages adoptés pour le développement du site afin de préserver au maximum la performance et l'expérience utilisateur, c'est donc logiquement le PHP (acronyme récursif pour "PHP : Hypertext Preprocessor") qui sera utilisé pour collecter, traiter et restituer les données. Les actions réalisées par les visiteurs sur le site seront mesurées via du CSS (Cascading Style Sheets). Enfin, grâce à un paramètre d'URL, il sera possible d'identifier un visiteur et son parcours sur plusieurs pages du site. Les données collectées seront enregistrées dans une base de données.

3.2.1 Collecte des données en PHP

Le PHP est un langage qui permet de construire des pages contenant du code HTML côté serveur et de restituer le résultat dans le navigateur côté client. C'est un avantage important pour les éditeurs de site. Ils peuvent ainsi exécuter du code avant l'affichage du contenu HTML et non accessible pour un utilisateur depuis le navigateur. Cette méthode permet d'assurer une collecte des données à chaque page. A l'inverse, le code JavaScript peut être problématique dans certains cas. S'il est appelé tardivement sur la page, un utilisateur peut quitter le site avant même que le script ait eu le temps de s'exécuter. De plus, beaucoup de fonctions JavaScript ne sont pas compatibles sur les anciens navigateurs. Ce qui impose aux développeurs de prévoir et de tester son code à travers des versions différentes de navigateurs.

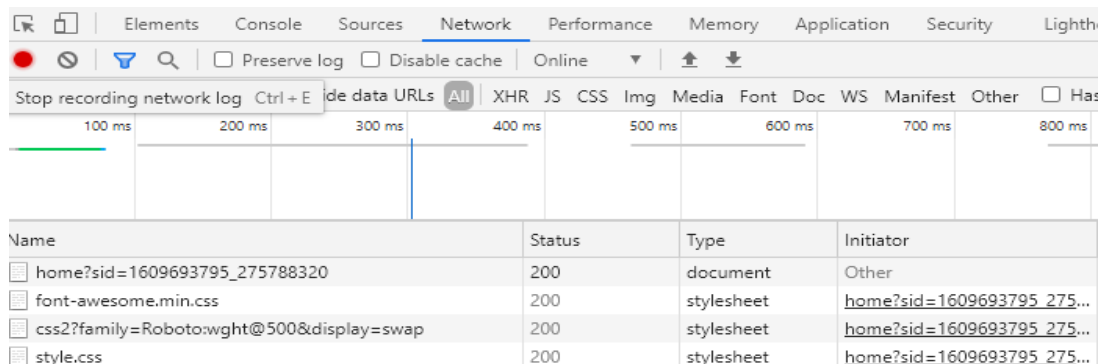
De ce fait, chaque page du site www.guillaume-sinnaeve.com est développée en PHP. Au début de chacune d'entre elles est inséré l'appel à un fichier nommé collect.php¹⁰. Ce dernier contient toute la mécanique qui va permettre de calculer la valeur de l'identifiant visiteur et de collecter les différentes informations concernant l'utilisateur, les pages et les interactions effectuées sur le site. A la fin de ce fichier, toutes les données récupérées sont enregistrées dans une base de données.

```
<?php
    include '../analytics/collect.php';
?>

<!DOCTYPE html>
<html>
```

Exemple d'une insertion du fichier collect.php sur une des pages du site

Le PHP envoie les informations du côté du serveur. Les données transmises ne sont donc pas visibles par l'utilisateur. En regardant les ressources chargées dans le *Network* du navigateur, seul l'appel au fichier apparaît.



Name	Status	Type	Initiator
home?sid=1609693795_275788320	200	document	Other
font-awesome.min.css	200	stylesheet	home?sid=1609693795_275...
css2?family=Roboto:wght@500&display=swap	200	stylesheet	home?sid=1609693795_275...
style.css	200	stylesheet	home?sid=1609693795_275...

Visibilité de la ressource home.php dans le Network sur la page <https://www.guillaume-sinnaeve.com/fr/home>

¹⁰ Page collect.php : <https://github.com/gsinna/master-miashs/blob/master/site/analytics/collect.php>

Ce qui peut être problématique pour l'éditeur de site qui met en place cette solution et qui souhaite vérifier la pertinence de ce qu'il collecte. Pour pouvoir effectuer ce contrôle de qualité, il est possible d'utiliser les en-têtes de réponse du fichier PHP chargé sur la page. Dans le cadre de l'expérience, les en-têtes sont préfixées par le terme "analytics_" et concaténées avec l'information collectée (exemple : analytics_device_browser: Chrome).

▼ Response Headers

```
analytics_date_day: Thursday
analytics_date_full: 12/17/2020
analytics_date_hour: 23
analytics_date_minute: 53
analytics_device_browser: Chrome
analytics_device_category: Desktop
analytics_device_language: fr
analytics_device_system: Windows
analytics_device_useragent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
like Gecko) Chrome/87.0.4280.101 Safari/537.36
analytics_event_custom_1
analytics_event_custom_2
analytics_event_name: page_view
analytics_geo_city
analytics_geo_country
analytics_geo_region
analytics_page_uri: /fr/home
analytics_referrer_full: https://www.guillaume-sinnaeve.com/fr/analyzes?sid=1608127638_757487712
analytics_referrer_host: www.guillaume-sinnaeve.com
analytics_referrer_path: /fr/analyzes
analytics_request_time: 1608245605289
analytics_session_id: 1608245605_695853376
```

Exemple d'en-têtes de réponse de la page <https://www.guillaume-sinnaeve.com/fr/home>

Le PHP permet donc de collecter les informations des pages. Néanmoins, il est intéressant de connaître les interactions des utilisateurs au sein d'une page pour mesurer leur engagement avec le site. Ne pouvant pas utiliser le JavaScript, nous allons voir comment le CSS peut répondre à notre besoin.

3.2.2 Mesure des interactions en CSS

Savoir si une page est vue par les utilisateurs est une première étape. Mais cette information seule n'est pas suffisante pour juger de la pertinence du contenu sur la page. En JavaScript, il est possible de mesurer la progression de lecture du contenu et les actions réalisées sur les pages, comme les clics sur les boutons. Des fonctions d'écoute (en anglais *addEventListener*) permettent d'exécuter du code suite à l'action d'un utilisateur, comme le scroll et le clic, dans l'objectif d'envoyer des informations à une solution d'analyse, par exemple. Néanmoins, dans le cadre de cette expérience, le JavaScript est prohibé. Par conséquent, nous allons utiliser un autre langage.

Le CSS permet de mettre en forme le contenu HTML. Il définit les propriétés des éléments de la page telles que la couleur, la taille, l'opacité, la position, etc. Un des avantages du CSS est qu'il permet de définir des propriétés différentes en fonction d'un état spécifique d'un élément grâce aux pseudo-classes. Par exemple, l'état initial d'un bouton peut avoir une couleur de fond bleue et devenir verte au survol de la souris par l'utilisateur, via la pseudo-class *hover*.

Pour l'expérience, des boutons de type chevron sont positionnés sur différentes pages et permettent d'afficher du contenu au clic de l'utilisateur. Ils sont liés à des balises input de type "checkbox". Les pseudo-classes CSS *focus* et *checked* vont permettre de définir des propriétés CSS différentes pour ces états. Et grâce à la propriété *background-image*, qui permet de faire appel à une URL pour charger une image en arrière-plan, nous allons pouvoir appeler le fichier `collect.php` et mesurer les interactions des utilisateurs. En ajoutant des paramètres d'URL à ce fichier, il est possible de qualifier l'action effectuée par le visiteur et de rappeler le fichier `collect.php`. En revanche, une fois l'URL du fichier mise en cache dans le navigateur de l'internaute, elle n'est pas appelée à nouveau si un second clic est réalisé sur la même page. Ainsi, il est possible d'obtenir un ratio entre affichage de la page et clic unique d'un bouton permettant d'en connaître sa performance.

```
<style type="text/css">
  .call-me:focus {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=call_me');
    background-size: 0px;
  }
  #tms:checked + h2 {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=show_competences&event_custom_1=TMS');
    background-size: 0px;
  }
  #analytics:checked + h2 {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=show_competences&event_custom_1=Analytics');
    background-size: 0px;
  }
  #gcp:checked + h2 {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=show_competences&event_custom_1=GCP');
    background-size: 0px;
  }
  #apis:checked + h2 {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=show_competences&event_custom_1=Google APIs');
    background-size: 0px;
  }
  #others:checked + h2 {
    background-image: url('../analytics/collect?sid=<?php echo $session_id ?>&event_name=show_competences&event_custom_1=Others');
    background-size: 0px;
  }
</style>
```

Styles CSS appliqués aux boutons de la page <https://www.guillaume-sinnaeve.com/fr/home>

Nous savons désormais comment mesurer les pages vues et les interactions effectuées par les utilisateurs sur le site. Cependant, aucun identifiant ne permet, pour le moment, de connaître le parcours effectué par un visiteur spécifique. Les solutions d'analyse et média utilisent en général un cookie pour stocker un identifiant et faire persister l'information de page en page. Néanmoins, les cookies étant prohibés pour cette expérience, nous utiliserons un paramètre d'URL pour identifier les visiteurs.

3.2.3 Identifiant visiteur en paramètre d'URL

En ce qui concerne l'identifiant d'un utilisateur non authentifié sur un outil d'analyse, un cookie first est généralement la solution envisagée. Par exemple avec Google Analytics, un cookie nommé “_ga” contient un identifiant permettant l'attribution des pages vues et des interactions réalisées à un même utilisateur. Nous avons vu précédemment que les cookies ont la vie dure actuellement et dans le cadre de l'expérience, ils sont bannis. Pour suivre l'utilisateur sur le site www.guillaume-sinnaeve.com, chaque URL de page et d'événement contient un paramètre d'URL ayant pour valeur un identifiant unique lié à l'internaute qui visite le site.

Si un utilisateur se rend sur le site, un paramètre “sid” sera présent dans l'URL. La valeur correspond à une concaténation entre un *timestamp* à 10 chiffres et un numéro aléatoire à 10 chiffres. Un *underscore* permet de faire la liaison entre les deux (exemple : sid=1608305029_984833821). Ce paramètre est inséré après chaque URL de redirection interne du site. Cette technique permet d'avoir une valeur unique pour chaque visiteur du site. Un des avantages de cette solution est qu'aucune information n'est stockée sur l'appareil de l'utilisateur. En revanche, un des inconvénients est la facilité d'accès pour un utilisateur. Il peut modifier la valeur du paramètre. Par conséquent, l'ajout d'une sécurité semble nécessaire. Si la valeur ne respecte pas le format attendu ou si le paramètre est supprimé par l'utilisateur, alors le script PHP réattribue un nouvel identifiant au bon format.

```
// Set max inactivity session duration in seconds
$max_session_duration = 1800;

// Get the value of session ID
$session_id = isset($_GET['sid']) ? $_GET['sid'] : '';

// Flag to block data request
$data_request = true;

// session ID is defined
if($session_id != '') {
    // session ID has not the format expected
    if(preg_match('/[0-9]{10}_[0-9]{10}/', $session_id)) {
        $session_id = round(microtime(true)) . '_' . mt_rand(0000000000,(int) 9999999999);
        $data_request = false;
        header('location: ' . preg_replace('~(\\?&)sid=[^&]*~', '$1', $_SERVER['REQUEST_URI']) . 'sid=' . $session_id);
    }
    else {
        include 'connection.php';

        $sql = "SELECT 'session_id', MAX('timestamp') as lastTimestamp FROM 'data' WHERE 'session_id' = '" . $session_id . "'";
        $res = mysqli_query($conn, $sql);
        mysqli_close($conn);
        $data = [];

        while($row = $res->fetch_assoc()) {
            array_push($data, ["session_id" => $row['session_id'], "lastTimestamp" => $row['lastTimestamp']]);
        }

        if(isset($data[0]['session_id']) && isset($data[0]['lastTimestamp'])) {
            $dif = strtotime(date("Y-m-d H:i:s")) - strtotime($data[0]['lastTimestamp']);

            //If the last session ID interaction is greater than 30 min, we change session_id
            if($dif > $max_session_duration) {
                header('location: ' . preg_replace('~(\\?&)sid=[^&]*~', '$1', $_SERVER['REQUEST_URI']) . 'sid=' . round(microtime(true)) . '_' . mt_rand(0000000000,(int) 9999999999));
                $data_request = false;
            }
        }
    }
}
}
```

Calculs effectués pour attribuer une valeur au paramètre “sid” de l'URL

Dans notre expérience, nous ne pouvons pas utiliser le terme “utilisateur” au sein de l'outil d'analyse car l'identifiant est réinitialisé à chaque fois qu'un internaute visite le site. Nous utilisons donc le terme de “visiteur unique”. Les notions de nouveaux visiteurs et de visiteurs de retour n'existent pas. Pour que le comportement de cet identifiant corresponde le plus à ceux proposés par les autres outils d'analyse, une règle est mise en place. Si un utilisateur est inactif pendant plus de 30 minutes, c'est-à-dire qu'il ne génère aucune requête de page ou

d'événement, alors un nouvel identifiant est généré à la prochaine requête. Cela permet d'éviter un calcul de temps de visite faussé par des utilisateurs qui reviendraient sur le site avec le même identifiant, grâce à un lien dans la barre des favoris par exemple.

Nous avons détaillé précédemment les solutions techniques utilisées dans le cadre de l'expérimentation. Le PHP pour la collecte côté serveur, le CSS pour mesurer les interactions et un identifiant dans un paramètre de l'URL pour suivre le visiteur nous permettent de respecter les règles établies pour ce projet. Nous pouvons maintenant sélectionner, traiter, stocker et restituer avec des schémas graphiques les données collectées sur les utilisateurs du site.

3.3 De la collecte à la visualisation des données

Maintenant que nous connaissons les pré-requis législatifs à respecter et les contraintes techniques imposées pour cette expérience, nous pouvons procéder à la collecte et au traitement des données. Dans notre exemple (mais cela vaut en général), il n'est pas souhaitable de collecter toutes les données à notre disposition car nous ne les utiliserions pas effectivement. C'est à ce moment précis que se pose la question de la finalité des données. Chaque information récupérée doit répondre à un besoin défini. Dans un premier temps, nous sélectionnerons les données à collecter pour mesurer l'audience du site. Ensuite, nous effectuerons un traitement sur celles-ci pour les regrouper, dans les cas où il est possible de le faire, et simplifier le stockage dans la base de données. Enfin, nous mettrons en place sur la page d'analyse du site www.guillaume-sinnaeve.com une restitution graphique des données collectées.

3.3.1 Connaître ses visiteurs

Comme nous l'avons vu précédemment dans la première partie, mesurer son audience est une des finalités de la collecte des données. Pour identifier le profil des internautes visitant le site internet, nous utiliserons les informations relatives à leur localisation, à leur appareil et à leur source de trafic.

Dans le cadre de l'exemption CNIL, la géolocalisation "ne doit pas être plus précise que l'échelle de la ville"¹¹. Grâce à l'adresse IP de l'utilisateur et à l'API proposé par ip-api.com¹², nous récupérerons les informations liées à son pays, sa région et sa ville. En complétant avec la langue du navigateur, toutes ces données permettent d'adapter les traductions du site en fonction des pays mais également de connaître les lieux où se situent les opportunités éventuelles, dans le cadre d'une mise en relation professionnelle.

¹¹ La géolocalisation ne doit pas être plus précise que l'échelle de la ville : https://www.cnil.fr/sites/default/files/atoms/files/projet_de_recommandation_cookies_et_autres_traceurs.pdf

¹² API utilisée pour la géolocalisation : <http://ip-api.com>

```
// Get visitor IP
$visitor_ip = $_SERVER['REMOTE_ADDR'];
// Connect to ip-api.com and get values
$ip_query = @unserialize(file_get_contents('http://ip-api.com/php/' . $visitor_ip));
// Get informations only if the request is OK
if($ip_query && $ip_query['status'] == 'success')
{
    //-----
    // geo_country
    //-----

    $geo_country = $ip_query['country'];

    //-----
    // geo_region
    //-----

    $geo_region = $ip_query['regionName'];

    //-----
    // geo_city
    //-----

    $geo_city = $ip_query['city'];
}
```

Récupération de l'adresse IP pour déterminer le pays, la région et la ville de l'utilisateur

Les données relatives à l'appareil peuvent être calculées à partir de l'agent utilisateur (ou *user agent* en anglais). Ce dernier est composé d'une chaîne de caractère offrant la possibilité à un serveur d'identifier un profil d'appareil. Plusieurs informations peuvent être obtenues par cette méthode comme la catégorie de l'appareil (ordinateur de bureau, tablette, mobile, etc.), le système d'exploitation (Windows, iOS, Android, etc.) et le nom du navigateur (Chrome, Safari, Firefox, etc.). Pour l'expérience, nous collectons ces données pour connaître l'audience mais également pour vérifier si le site est bien fonctionnel avec la configuration de l'utilisateur et de l'optimiser dans le cas contraire.

Dans notre cas, les sources de trafic font référence aux sites qui possèdent un lien redirigeant vers le nom de domaine de l'expérience. Cette information peut se récupérer depuis l'URL de la page précédente et est utile pour connaître la notoriété du site sur le Web. Si cette valeur est vide, c'est que l'utilisateur est venu en direct, en inscrivant l'URL dans son navigateur ou via un favori. Nous enregistrons le referrer pour notre expérience. Dans l'objectif de tester différentes sources de trafic et acquérir des données, plusieurs campagnes sur les réseaux sociaux ont été mises en place à des dates différentes (voir Annexe 7).

Les données relatives aux appareils et aux sources de trafic nous en disent davantage sur les visiteurs du site. Et pour mieux connaître cette audience, les informations liées aux pages vues et aux événements effectués sur le site nous sont utiles. Nous allons donc sélectionner les données comportementales à collecter.

3.3.2 Le comportement sur le site

Les internautes interagissent avec le site internet, en chargeant des pages et en réalisant des actions comme les clics sur les boutons. Ce comportement peut être mesuré de différentes manières : le temps de chargement des pages, les pages visitées, les événements (actions réalisées sur une page) et les durées des visites.

Récupérer les informations de temps de chargement des pages aide à améliorer la performance du site Web, surtout quand elles sont croisées avec les données d'audience. Une page peut être plus longue à charger sur un téléphone mobile par exemple. Dans notre cas, il n'est pas nécessaire de collecter cette information car le site est de taille modeste et ne charge que peu de scripts.

Dans l'objectif d'identifier les pages vues par les visiteurs du site, nous devons collecter l'URL concernée, et plus précisément le chemin qui la constitue, par exemple `/fr/home`. Il n'est pas nécessaire de collecter le nom de domaine car l'expérience est réalisée uniquement www.guillaume-sinnaeve.com. Cet élément de l'URL est également utile pour connaître sur quelles pages les événements sont effectués.

Pour distinguer si le comportement de l'utilisateur est lié à une page ou à un événement, nous allons alimenter une variable que nous enregistrerons dans la base de données. Elle aura pour valeur soit `page_view` pour une page, soit une valeur différente pour toute autre action (exemple : `call_me` pour la prise de contact téléphonique, `show_compétences` pour afficher les compétences sur la page d'accueil, etc.). Cette variable seule n'est pas suffisante pour certaines actions. En la combinant avec d'autres éléments, nous pouvons qualifier l'événement (par exemple indiquer la compétence vue dans une variable lors d'une action `show_compétences`).


```
//-----  
// event_name  
//-----  
$event_name = isset($_GET['event_name']) ? $_GET['event_name'] : 'page_view';  
  
//-----  
// event_custom_1  
//-----  
$event_custom_1 = isset($_GET['event_custom_1']) ? $_GET['event_custom_1'] : '';  
  
//-----  
// event_custom_2  
//-----  
$event_custom_2 = isset($_GET['event_custom_2']) ? $_GET['event_custom_2'] : '';
```

Récupération du nom de l'événement et des paramètres qui le qualifie via l'URL d'appel au fichier collect.php

Nous avons détaillé les données relatives aux visiteurs, et à leur comportement sur le site, qui seront collectées pour cette expérimentation. Certaines de ces informations, comme l'agent utilisateur, nécessitent un traitement avant de les stocker pour obtenir le système d'exploitation ou le nom du navigateur par exemple. Plusieurs actions et calculs sont requis. Nous allons les aborder dans la partie suivante.

3.3.3 Traitement et stockage

Nous savons désormais quelles informations nous souhaitons collecter sur le site et leur finalité. En amont du stockage dans la base de données, il peut être nécessaire d'effectuer un traitement. Il peut s'agir d'un nettoyage, d'un regroupement, d'un calcul ou d'un complément d'information sur les valeurs récupérées.

Une des problématiques fréquemment rencontrée par les outils analytiques est la génération de données provenant des bots informatiques. Ces programmes parasites viennent fausser la lecture des statistiques collectées sur le nombre de visiteurs, de pages vues et d'événements. Nous ne souhaitons pas que cela se produise dans le cadre de l'expérience et grâce à l'agent utilisateur, nous pouvons identifier une partie de ce trafic. Dans une majorité des cas, la chaîne de caractère contient le mot "bot". Si nous trouvons cette occurrence, alors nous empêchons l'enregistrement en base de données. Cette règle peut être améliorée et renforcée au cours de la vie du site pour bloquer les bots non couverts actuellement.

```
// Get the user agent
$device_userAgent = $_SERVER['HTTP_USER_AGENT'];

// Exclude Bot based on user agent
if(preg_match('/Bot|facebookexternalhit|PostmanRuntime|Scoop\.it|Go-http-client/i', $device_userAgent)) {
    $data_request = false;
}
```

Condition pour exclure les bots de la collecte de données

Un regroupement peut être nécessaire, dans des cas spécifiques, pour éviter une duplication de lignes dans la base de données et complexifier la lecture. Pour l'expérience, nous collectons le referrer. Et lors des campagnes lancées sur Facebook (voir Annexe 7), plusieurs valeurs sont remontées pour cette même source de trafic : www.facebook.com, m.facebook.com et l.facebook.com. N'ayant pas un besoin d'analyse lié à cette distinction, une règle dans le fichier collect.php réunit ces 3 domaines sous la valeur www.facebook.com avant de la stocker.

```
// By default, referrer_full, referrer_host and referrer_path equal 'unknown'
$referrer_full = 'unknown';
$referrer_host = 'unknown';
$referrer_path = 'unknown';

// Check if http referer is not undefined
if(isset($_SERVER['HTTP_REFERER'])) {

    $referrer_full = $_SERVER['HTTP_REFERER'];

    if (true === strpos($_SERVER['HTTP_REFERER'], 'guillaume-sinnaeve.com')){
        $referrer_host = 'www.guillaume-sinnaeve.com';
    }
    else if (true === strpos($_SERVER['HTTP_REFERER'], 'facebook')){
        $referrer_host = 'www.facebook.com';
    }
    else if ($_SERVER['HTTP_REFERER'] == 'android-app://com.linkedin.android/'){
        $referrer_host = 'www.linkedin.com';
    }
    else {
        $referrer_host = parse_url($_SERVER['HTTP_REFERER'], PHP_URL_HOST);
    }

    $referrer_path = parse_url($_SERVER['HTTP_REFERER'], PHP_URL_PATH);
}
else if(preg_match('/LinkedInApp/i', $device_userAgent)) {
    $referrer_full = 'https://www.linkedin.com/';
    $referrer_host = 'www.linkedin.com';
    $referrer_path = '/';
}
else {
    // Direct
    $referrer_full = 'Direct';
    $referrer_host = 'Direct';
    $referrer_path = 'Direct';
}
```

Calculs effectués pour déterminer les valeurs liées aux referrers

Dès lors que les données non désirées sont écartées, nous pouvons procéder aux calculs des valeurs que nous souhaitons obtenir à partir des informations collectées. Ce qui est le cas, dans le cadre de l'expérience, de la catégorie de l'appareil, du système d'exploitation et du nom du navigateur. Ci-après, un exemple d'une valeur de l'agent utilisateur obtenue sur le navigateur Safari exécuté sur un iPhone X :

Mozilla/5.0 (iPhone; CPU iPhone OS 13_2_3 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/13.0.3 Mobile/15E148 Safari/604.1

Grâce aux termes iPhone et Mobile, nous pouvons attribuer les valeurs "Mobile" à la catégorie d'appareil et "iOS" au système d'exploitation. Concernant le nom du navigateur, la présence du terme Safari et les autres informations en notre possession nous permettent de dire que Safari est bien le nom du navigateur utilisé.

```
// By default, device_system equals 'unknown'
$device_system = 'unknown';

// iOS
if(preg_match('/iPhone|iPad|iPod/i', $device_userAgent)) {
    $device_system = 'iOS';
}

// Mac OS
else if(preg_match('/Mac/i', $device_userAgent)) {
    $device_system = 'Mac OS';
}

// Windows
else if(preg_match('/Win/i', $device_userAgent)) {
    $device_system = 'Windows';
}

// Android
else if(preg_match('/Android/i', $device_userAgent)) {
    $device_system = 'Android';
}

// UNIX
else if(preg_match('/X11/i', $device_userAgent)) {
    $device_system = 'UNIX';
}

// Linux
else if(preg_match('/Linux/i', $device_userAgent)) {
    $device_system = 'Linux';
}

// BlackBerry
else if(preg_match('/BB10|PlayBook/i', $device_userAgent)) {
    $device_system = 'BlackBerry';
}
```

Calculs effectués pour déterminer le système d'exploitation de l'utilisateur

Avant de procéder à l'enregistrement et au stockage des informations, d'autres propriétés sont utiles comme le *timestamp*. Il permet de suivre l'ordre du parcours effectué par le visiteur, la durée de sa visite et également d'établir des rapports de fréquences. Avec le *timestamp*, nous obtenons la date (année, mois, jour de la semaine), le numéro du jour, l'heure et la minute

auxquels la requête a été transmise. Dès lors que tout est prêt, nous pouvons exécuter une instruction en PHP et en SQL (Structured Query Language) pour enregistrer les informations dans la base de données (voir Annexe 8 pour le schéma de la base de données).

Les données sont maintenant traitées et stockées dans la base de données. Nous pouvons désormais les utiliser au sein d'un outil de visualisation graphique pour faciliter la lecture. Dans la partie suivante, nous utiliserons une des pages du site pour afficher des rapports d'analyse sur les visiteurs, les pages et les événements du site.

3.3.4 Représentation graphique des données

Nous avons mis en place un outil de collecte et de stockage de données à des fins de mesure d'audience. Nous pourrions conclure ici et utiliser la base de données avec des requêtes SQL pour lire les indicateurs de performance qui nous intéressent car beaucoup d'outils permettent d'avoir accès à ces données sans passer par une interface de rendu. Cependant, la lecture des informations est complexe et ne permet pas de comprendre rapidement les résultats obtenus. Par conséquent, nous allons construire une page d'analyses¹³ sur le site contenant plusieurs représentations graphiques des données collectées (voir Annexe 9). Et par souhait de transparence avec les visiteurs du site et les données qui les concernent, cette page est accessible par tous.

Pour une mise en forme agréable et interactive, nous allons utiliser les librairies C3¹⁴ et D3¹⁵ de JavaScript. Elles permettent de créer facilement, en plus des tableaux, des graphiques en bâton, en ligne ou encore en donut à partir d'une plage de données. Les fichiers requests.php¹⁶ et dash.js¹⁷ contiennent la construction des différents graphiques proposés sur la page. Cette dernière est composée de 4 parties consacrées aux données en temps réel, relatives aux visiteurs, aux pages et aux événements.

La partie intitulée Temps réels va contenir les données des visiteurs du site présents dans les 30 dernières minutes (voir Annexe 9). Elle permet un aperçu rapide de l'audience du site et un contrôle de qualité sur la remontée d'informations liés aux visiteurs, aux pages et aux événements. Des graphiques issus des trois parties, nommées Visiteurs, Pages et Événements, composent ce bloc et mettent en avant quelques indicateurs de performance clés avec des niveaux de lecture différents.

Le bloc Visiteurs affiche les informations relatives aux internautes et à leur visite sur les 25 derniers mois (voir Annexe 9). Des indicateurs affichent le nombre total de visiteurs, le taux de rebond, la durée moyenne des visites ou encore le nombre de visites avec des événements.

¹³ Pages d'analyse : <https://www.guillaume-sinnaeve.com/fr/analyzes> et <https://www.guillaume-sinnaeve.com/en/analyzes>

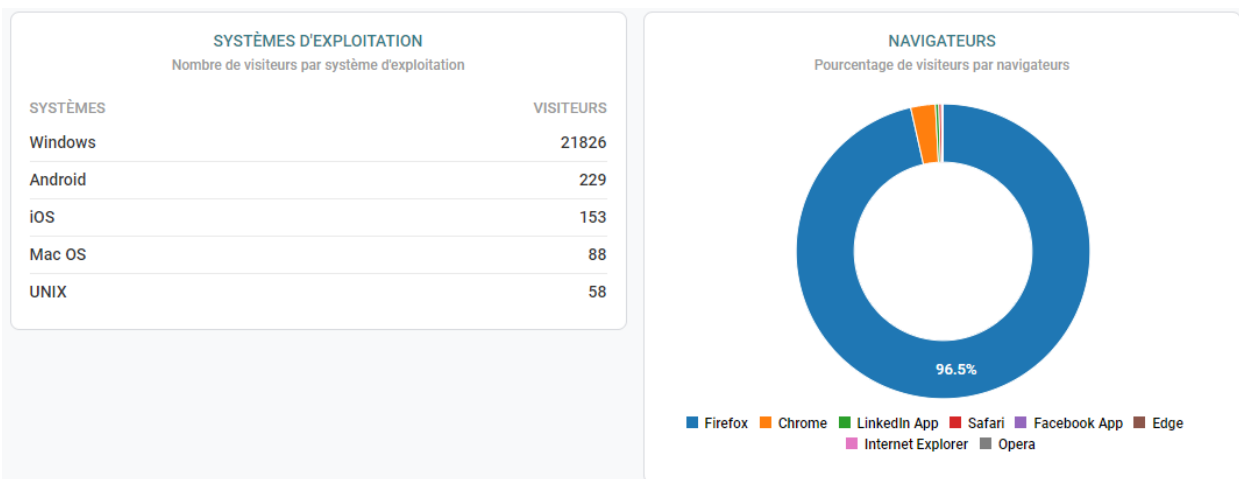
¹⁴ Librairie C3.js : <https://c3js.org/>

¹⁵ Librairie D3.js : <https://d3js.org/>

¹⁶ Page requests.php : <https://github.com/gsinna/master-miashs/blob/master/site/analytics/requests.php>

¹⁷ Page dash.js : <https://github.com/gsinna/master-miashs/blob/master/site/analytics/dash.js>

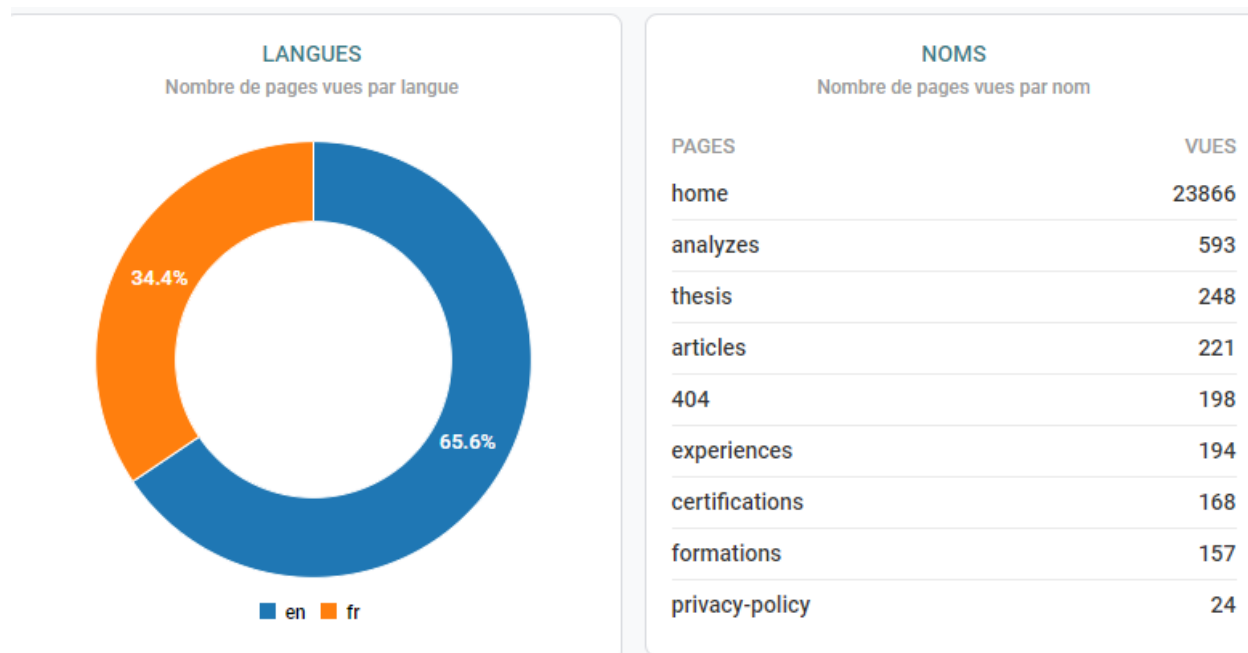
Ces quatre résultats offrent une vision globale des visiteurs tandis que les autres rapports de cette partie apportent davantage de précisions. Nous pouvons analyser les visiteurs par fréquence de visites, pour optimiser les campagnes et les mises à jour de contenu, par configuration d'appareil, pour vérifier que le site est fonctionnel sur différents navigateurs, ou encore par villes, pour identifier les secteurs où des opportunités professionnelles seraient envisageables.



Rapports liés aux systèmes d'exploitations et aux navigateurs des visiteurs

Nous pouvons constater sur les rapports ci-dessus que les visiteurs possèdent majoritairement un appareil avec Windows comme système d'exploitation et naviguent sur le site avec le navigateur Firefox.

Ensuite, la section Pages met en avant les données correspondant aux pages vues par les utilisateurs durant leur visite sur les 25 derniers mois (voir Annexe 9). Trois indicateurs clés de performances sont mis en avant ici : le nombre total de pages vues, la moyenne de pages vues par visiteur et le nombre total de pages d'erreur ou non trouvées (404) vues. Les autres rapports de cette section proposent une analyse par fréquence, par page d'arrivée et de sortie sur le site, par langue des pages et par nom de page.



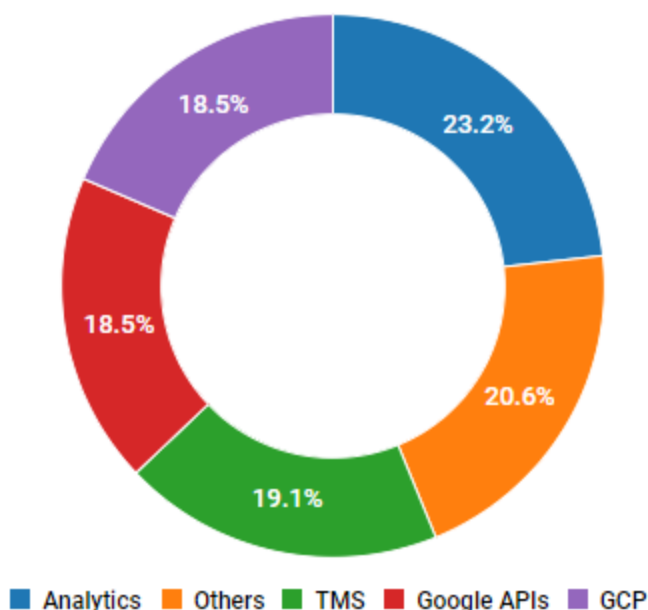
Rapports liés aux langues et aux noms des pages vues

Le rapport précédent met en évidence le pourcentage de pages vues par langue et le nombre de pages vues par nom. On constate que le site est parcouru essentiellement en anglais et que la page d'accueil est trente fois plus consultée que la page suivante.

Enfin, la partie Événements reflète les actions effectuées par les internautes durant leur visite sur les 25 derniers mois (voir Annexe 9). Les trois chiffres importants mis en avant dans cette partie sont le nombre total d'événements, le nombre de visiteurs ayant lu un article et le nombre de visiteurs ayant cliqué pour initier un appel téléphonique. Les autres rapports de cette partie proposent une analyse par fréquence et par type d'événement effectué comme la consultation des compétences, du parcours professionnel ou encore des certifications.

COMPÉTENCES RECHERCHÉES

Nombre d'événements par compétence



Rapport lié aux compétences consultées par les visiteurs

Le graphique ci-dessus indique le nombre d'événements réalisés par compétence. D'après les résultats obtenus, nous pouvons en déduire que toutes les compétences sont consultées à parts quasi égales même si la compétence Analytics se détache légèrement.

La page d'analyse du site propose une mise en relation des données collectées sous forme de plusieurs représentations graphiques. Cette page peut être complétée par d'autres rapports comme, par exemple, un flux de parcours réalisés depuis la page d'accueil. Les rapports doivent refléter le besoin d'analyse émis par l'équipe métier d'une entreprise. Dans le cadre de l'expérience, certains rapports affichent les valeurs "unknown". Elles indiquent que des calculs, comme le nom du navigateur ou le système d'exploitation, n'a pas pu trouver de résultat. Par conséquent, une analyse supplémentaire doit être menée au sein de ma base de données pour comprendre la raison, en utilisant l'agent utilisateur, et optimiser le calcul pour une meilleure collecte des données.

Conclusion

La collecte des données est essentielle pour les entreprises à des nombreux titres : connaître leur audience, optimiser la conversion ou encore monétiser leur contenu. Grâce à l'exploitation des cookies et d'une faille de sécurité, les solutions d'analyses, médias et autres, ont créé un modèle économique mettant en danger l'aspect privé des données personnelles des internautes. C'est pourquoi le RGPD, les navigateurs et certaines extensions bloquant les publicités ont choisi de se positionner pour le respect de la vie privée des utilisateurs. Face à ces nombreux changements et aux obligations législatives qui durcissent, les entreprises ont été contraintes d'adapter leur organisation juridique et d'investir du temps de développement pour s'assurer une mise en conformité légale au plus tard à la date butoire du 31 mars 2021.

Bien entendu, les entreprises souhaitent conserver une qualité de service équivalente à l'avant RGPD. Comprendre le parcours des utilisateurs et leur comportement sur le site grâce à la collecte de données reste nécessaire pour proposer un contenu adapté à son audience. Il n'y a, à ce jour, pas de solution parfaite pour répondre à toutes les attentes des entreprises. Par conséquent, les éditeurs de site doivent composer avec ce qui existe actuellement sur le marché (exemption CNIL, server-side, navigation authentifiée, *machine learning*) tout en proposant, dans la majorité des cas, un bandeau de consentement, s'exposant ainsi à perdre de précieuses informations sur leurs utilisateurs.

Nous avons nous-même expérimenté une solution conforme au RGPD, sur le site www.quillaume-sinnaeve.com, et fonctionnelle dans un potentiel monde futur sans cookies. En plus d'être sans JavaScript, la solution tire profit des technologies stables et renommées que sont le PHP et le CSS, utilisés par ailleurs à d'autres fins lors du développement du site. Nous sommes parvenus à récupérer, et à afficher dans des rapports graphiques, des données comportementales et relatives aux visiteurs tout en respectant les pré-requis législatifs dans le cadre de l'exemption CNIL. Il va sans dire, néanmoins, que cette solution est perfectible et se limite à un site ayant pour seule finalité la mesure d'audience. Pour des éditeurs de site plus complexes qui souhaiteraient faire de l'activation de données, il faudrait étudier les obligations légales et la technicité de l'outil proposé ici pour l'adapter aux différents besoins.

Les démarches de mise en conformité au RGPD sur le Web semblent bien engagées par les entreprises, même si certains textes de lois apparaissent encore flous et donc soumis à interprétation. Les autorités de protection des données des différents pays de l'Union Européenne continuent de contrôler et de sanctionner les manquements à la loi.

Pour les applications mobiles, où la notion de cookie n'existe pas, la collecte de données est également soumise au consentement de l'utilisateur et aux textes législatifs. Les éditeurs ne sont d'ailleurs pas en risque puisque, par exemple, Apple ajoute des contraintes techniques encore plus strictes pour garantir la protection de la vie privée de ses utilisateurs mobiles avec la version d'iOS 14 de janvier 2021. Les informations collectées sur les utilisateurs par l'application mobile doivent être affichées sur l'App Store et les développeurs devront demander

le consentement explicite pour l'IDFA (en anglais IDentifier For Advertisers ou identifiant pour les annonceurs) à l'ouverture de l'application. Cet identifiant est très utilisé côté mobile et s'apparente à un cookie lié à un utilisateur côté Web. Des impacts non négligeables sont donc à prévoir pour l'exploitation des données à des fins d'activation et d'optimisation média.

Aujourd'hui, il existe des modules de consentement (CMP) pour répondre aux obligations législatives mais ils ne couvrent pas les contraintes techniques imposées par Apple. Plusieurs solutions sont à l'étude pour permettre aux différents éditeurs d'applications et aux acteurs publicitaires et technologiques de continuer à collecter et activer la donnée tout en respectant la volonté d'Apple de laisser à ses utilisateurs le choix de l'exploitation de leurs données.

En résumé, si la collecte des données de leurs utilisateurs est essentielle pour les entreprises, force est de constater que les dernières années ont vu leur lot d'abus et de scandales de la part de nombreux acteurs principaux du marché. Il est nécessaire aujourd'hui de poser des bases juridiques à la collecte et à l'utilisation des données personnelles. La mort du cookie et l'exigence d'un consentement impliquent une baisse non négligeable de la qualité des données exploitables par les sites internet et les solutions publicitaires. Néanmoins, il semble inconcevable que les entreprises ne trouvent pas des solutions satisfaisantes pour tous et qui répondent à ces nouvelles exigences. Reste à savoir désormais quelles formes elles pourront prendre.

Annexes

Annexe 1 - Questionnaire - Impact du RGPD sur Entreprise A

1. Quel est l'intitulé de votre poste et vos missions au sein de votre entreprise ?

Je suis Web Analytics Manager. Mes missions consistent à collecter l'ensemble des données digitales permettant d'optimiser nos produits (applications, sites web) et nos campagnes marketing. Cela passe par la mise en place de plans de marquage, de recettes techniques, de tableaux de bord et de reporting, d'études, d'activation (retargeting, ciblage, personnalisation) et de test A/B.

2. Comment suivez-vous les mises à jour des recommandations émises par la CNIL et les DPAs des autres pays (dans le cadre où votre entreprise est internationale) sur le RGPD ?

Nos équipes juridiques nous tiennent informés de ces évolutions via des newsletter hebdomadaires et des Coproj bi-mensuel. Nos partenaires se rapprochent de nous quand ces évolutions ont un impact direct sur leurs solutions. Je me tiens également informé grâce aux abonnements à des newsletter de cabinets de conseil et d'agence média.

3. Avez-vous dû créer une équipe dédiée pour suivre la mise en conformité au RGPD ?

Nous avons mis en place un comité projet ayant lieu toutes les 2 semaines réunissant tous les acteurs impactés par la mise en conformité :

- équipe juridique
- équipe data (data science et web analyse)
- équipe front/SI
- équipe marketing

Cette réunion nous permet de faire le point sur l'avancée du projet et des dernières actions prises et à prendre pour arriver à notre mise en conformité cible.

4. Si une équipe au sein de votre entreprise existe, de quel département dépend-t-elle (Sécurité, DSI Marketing, Juridique, etc.) ?

L'équipe du chef de projet est rattachée au département juridique.

5. La CNIL a mis à jour ses recommandations en septembre 2020. Avez-vous anticipé la mise en place des nouvelles règles de la CNIL pour la collecte du consentement sur les cookies ?

Oui

6. Si oui, quels sont les dispositifs que vous avez mis en place ?

Nous avons mis en place un comité de projet.

7. Si non, quelles actions allez-vous mettre en place dans les prochains mois concernant la collecte du consentement pour les cookies ?

N/A

8. Pensez-vous que les différents acteurs (analytique, média, etc.) ont réussi à anticiper ces évolutions ?

Non car, en plus des nouvelles législations, il y a un mouvement de marché des plus gros acteurs vers un plus grand respect des données utilisateurs (ITP, fin des cookies tiers). Je n'ai pas l'impression que les éditeurs ont pu se réunir pour déterminer toutes les conséquences et solutions à l'ensemble de ces changements concomitants.

9. Traitez-vous le sujet de la collecte du consentement sur les traceurs à la fois pour vos sites web et vos applications mobiles ?

Oui

10. Avez-vous lancé des actions d'A/B testing de votre bandeau cookie et/ou de votre CMP afin d'optimiser la collecte du consentement ?

Oui

11. Quels sont les impacts du RGPD au niveau technique ?

Nous avons mis en place de nouveaux une CMP et fait évoluer des outils existants (tracking d'attribution différent).

12. Est-ce que le tracking server-side est une solution que vous avez envisagée ?

Non car concernant le RGPD qu'importe la méthodologie utilisée (client side ou server side) le consentement reste obligatoire.

13. Avez-vous mis en place des stratégies pour tout de même suivre les visiteurs qui ont refusé le tracking ?

Oui avec les solutions le permettant et qui sont acceptées et exemptées par la CNIL. A savoir AT Internet.

14. Au-delà des mises à jour techniques de vos outils (CMP, TMS) pour mieux collecter le consentement utilisateur sur les cookies/traceurs, avez-vous lancé un projet de réalignement de la gouvernance data au sein de votre entreprise ?

Oui

15. Quels impacts a eu le RGPD sur les données que vous utilisez pour piloter votre stratégie marketing ?

Nous anticipons de fortes pertes de données d'attribution. Nous devons donc passer de modèles déterministes à des modèles plus probabilistes et d'échantillonnage.

16. Quels sont les impacts sur vos investissements média ?

Pour le moment rien n'est acté nous attendons que notre mise en conformité soit totale pour voir l'impact sur la collecte et ainsi comprendre l'impact probable sur nos investissements média.

17. Prévoyez-vous ou avez-vous anticipé un éventuel report de budget média vers du SEO ou des partenariats ?

Oui

18. Est-ce que la mise en conformité au RGPD a engendré un coût budgétaire pour votre site ?

Oui car nous faisons appel à des consultants pour nous aider à gérer le projet (mise en place opérationnelle, conséquences éventuelles...) et nous implémentons de nouvelles solutions de CMP.

19. Identifiez-vous d'autres impacts que ceux évoqués précédemment concernant la mise en conformité du RGPD ?

Non

20. Pensez-vous que cette évolution de la législation changera drastiquement votre manière de travailler ?

Oui il y a fort à parier que de nombreux outils vont changer et d'autres vont émerger. Il faudra donc se former à ces nouvelles technologies.

Annexe 2 - Questionnaire - Impact du RGPD sur Entreprise B

1. Quel est l'intitulé de votre poste et vos missions au sein de votre entreprise ?

Global Analytics Manager

2. Comment suivez-vous les mises à jour des recommandations émises par la CNIL et les DPAs des autres pays (dans le cadre où votre entreprise est internationale) sur le RGPD?

Nous suivons les mises à jour des recommandations émises par la CNIL via des communications en interne et de la veille.

3. Avez-vous dû créer une équipe dédiée pour suivre la mise en conformité au RGPD ?

Un point de contact existe au sein du département digital, qui permet de centraliser tous les points et initiatives. En relais, dans les différents pays où notre groupe est présent, nous comptons sur des DPO.

4. Si une équipe au sein de votre entreprise existe, de quel département dépend-t-elle (Sécurité, DSI Marketing, Juridique, etc.) ?

CDO = Département Digital du groupe cross Brand et Pays

5. La CNIL a mis à jour ses recommandations en septembre 2020. Avez-vous anticipé la mise en place des nouvelles règles de la CNIL pour la collecte du consentement sur les cookies ?

Nous avons discuté en interne des nouvelles règles de la CNIL.

6. Si oui, quels sont les dispositifs que vous avez mis en place ?

N/A

7. Si non, quelles actions allez-vous mettre en place dans les prochains mois concernant la collecte du consentement pour les cookies ?

N/A

8. Pensez-vous que les différents acteurs (analytique, média, etc.) ont réussi à anticiper ces évolutions ?

Non pas totalement. La population digitale expert comprend les changements impulsés par la CNIL. Il y a des acteurs comme ceux qui sont à la base de la collecte et la partie analytique qui ont bien réussi mais en revanche les équipes médias ne comprennent pas forcément tous les tenants et les aboutissant. Notamment le retargeting ne fonctionnera plus par la suite.

9. Traitez-vous le sujet de la collecte du consentement sur les traceurs à la fois pour vos sites web et vos applications mobiles ?

Il y a peu d'applications au sein de notre groupe, mais la logique de traitement du consentement reste la même que pour les sites web.

10. Avez-vous lancé des actions d'A/B testing de votre bandeau cookie et/ou de votre CMP afin d'optimiser la collecte du consentement ?

Oui des initiatives ont été prises par quelques marques et pays pour tester la présence du bandeau, le message, les boutons. Pour répondre à la question comment perdre le moins de données via les obligations législatives.

11. Quels sont les impacts du RGPD au niveau technique ?

Notre partenaire est One Trust qui est configuré via GTM.

Les changements des modalités de la RGPD ont pour principaux impacts:

- La mise à jour des configurations au sein du JSON devant être intégrés sur chaque conteneur (un conteneur est égal à un pays / marque).
- La communication aux pays pour qu'ils téléchargent ce nouveau JSON au sein de leur conteneur (pour chaque site).

12. Est-ce que le tracking server-side est une solution que vous avez envisagée ?

Oui, par l'intermédiaire de GTM; nous avons discuté de ce sujet mais pas de grande avancée.

13. Avez-vous mis en place des stratégies pour tout de même suivre les visiteurs qui ont refusé le tracking ?

Non. Des questions se posent quant à la possibilité de réconcilier les données via les log de Cloudflare. Mais pour le moment rien ne se décide de ces discussions.

14. Au-delà des mises à jour techniques de vos outils (CMP, TMS) pour mieux collecter le consentement utilisateur sur les cookies/traceurs, avez-vous lancé un projet de réaligement de la gouvernance data au sein de votre entreprise ?

Le département CDO incarne cette centralisation de la gouvernance de la data, avec des équipes Media, Ecommerce, etc. qui ont pour but de développer des Guidelines pour le groupe et de le diffuser au sein du réseau.

15. Quels impacts a eu le RGPD sur les données que vous utilisez pour piloter votre stratégie marketing ?

Nous avons moins de données donc impact direct sur les initiatives media (audiences...), la connaissance client va être amoindrie. La création et la signification des audiences dans les outils vont être moindre, les programmes de fidélité et le CRM aussi.

16. Quels sont les impacts sur vos investissements média ?

Je ne sais pas exactement comment les équipes media travaillent sur ce point.

17. Prévoyez-vous ou avez-vous anticipé un éventuel report de budget média vers du SEO ou des partenariats ?

Je ne sais pas exactement comment les équipes media travaillent sur ce point.

18. Est-ce que la mise en conformité au RGPD a engendré un coût budgétaire pour votre site ?

Oui dans la mesure où si nous avons moins d'informations collectés via les cookies, il est dans notre intérêt de soigner la connaissance client en ajoutant des fonctionnalités comme la création de compte pour servir les programmes CRM, de fidélité, etc. Ce virage entraînerait des coûts supplémentaires dans le développement d'un site et au final dans les campagnes

19. Identifiez-vous d'autres impacts que ceux évoqués précédemment concernant la mise en conformité du RGPD ?

Oui, un très important, la compréhension des données à disposition par des profils non experts.

Si changement de Policy (ex: le DPO d'un pays décide de considérer Google Analytics comme un cookie de Performance et plus un cookie Strictement nécessaire), un impact "négligé" mais qui relève au final plus de la communication sera d'expliquer une perte de données qui ne résulte pas d'une baisse de trafic mais tout simplement d'un volume de visite pour lesquels l'internaute n'a pas donné son consentement.

"Moins 20%!!!" non vous ne faites pas -20% même si c'est dans l'absolu possible; via le changement de Policy, pour collecter de la donnée votre internaute doit donner son consentement ce qui n'était pas le cas avant.

20. Pensez-vous que cette évolution de la législation changera drastiquement votre manière de travailler ?

Oui, et ceci sous plusieurs angles:

- La formation: expliquer la nature des changements et leurs impacts (voir question 19)
- L'analyse: comprendre que le champs d'analyse se réduit
- L'activation: impact sur les stratégie media
- Le coût de développement d'un site: ajout de nouvelles fonctionnalités
- L'ADN d'une marque: programme de fidélité ou pas pour activer la connaissance client et légitimé la collecte d'information

Annexe 3 - Questionnaire - Impact du RGPD sur Entreprise C

1. Quel est l'intitulé de votre poste et vos missions au sein de votre entreprise ?

Responsable suivi e business et web analyse.

2. Comment suivez-vous les mises à jour des recommandations émises par la CNIL et les DPAs des autres pays (dans le cadre où votre entreprise est internationale) sur le RGPD ?

Les équipes directement concernées par l'application des recommandations de la CNIL car impactées par les obligations juridiques et réglementaires induites pour le bon exercice de leur activité exercent une veille constante des évolutions liées, en lien étroit avec les différentes solutions prestataires avec lesquelles elles travaillent, afin de coller au plus près des bonnes pratiques du marché.

3. Avez-vous dû créer une équipe dédiée pour suivre la mise en conformité au RGPD ?

Oui une équipe a été créée afin de notamment garantir en interne la bonne application des directives GDPR principales auprès des différentes équipes concernées.

4. Si une équipe au sein de votre entreprise existe, de quel département dépend-t-elle (Sécurité, DSI Marketing, Juridique, etc.) ?

La centralisation de ces sujets est affectée au département Big Data.

5. La CNIL a mis à jour ses recommandations en septembre 2020. Avez-vous anticipé la mise en place des nouvelles règles de la CNIL pour la collecte du consentement sur les cookies ?

Oui les nouvelles mesures de la CNIL ont été anticipées dans leurs grandes lignes depuis les précédentes directives.

6. Si oui, quels sont les dispositifs que vous avez mis en place ?

Ce sujet a principalement été anticipé par une réflexion autour du format du bandeau de collecte de privacy afin d'anticiper le dispositif permettant le meilleur compromis entre la collecte

du consentement d'une part et le respect des directives CNIL d'autre part, et observée via le déploiement d'ab tests sur notre site web.

7. Si non, quelles actions allez-vous mettre en place dans les prochains mois concernant la collecte du consentement pour les cookies ?

N/A

8. Pensez-vous que les différents acteurs (analytique, média, etc.) ont réussi à anticiper ces évolutions ?

Dans les grandes lignes oui car la vocation première du RGPD était claire dès le départ, des zones grises restent néanmoins à éclaircir, tant sur la latitude juridique que les acteurs peuvent conserver dans la mise en place de la collecte du consentement, que sur les solutions que les grands acteurs technologiques d'achat média et analytics vont déployer pour mieux répondre aux enjeux de sécurité de la donnée demain.

9. Traitez-vous le sujet de la collecte du consentement sur les traceurs à la fois pour vos sites web et vos applications mobiles ?

Le bandeau de collecte du consentement est actif sur l'ensemble de nos sites. Nous n'avons pas d'application mobile.

10. Avez-vous lancé des actions d'A/B testing de votre bandeau cookie et/ou de votre CMP afin d'optimiser la collecte du consentement ?

Oui, plusieurs ab test ont été lancés dans les derniers mois pour définir le bon format du bandeau.

11. Quels sont les impacts du RGPD au niveau technique ?

La RGPD impose un contrôle plus attentif quant à la qualité de la donnée collectée afin que celle-ci soit conforme aux principes d'anonymisation énoncés par celle-ci. Elle nécessite également de garantir les moyens de fournir à nos clients demandeurs une visibilité sur l'ensemble des données qui ont été collectées le concernant.

12. Est-ce que le tracking server-side est une solution que vous avez envisagée ?

C'est une solution que l'on envisage effectivement à l'heure actuelle.

13. Avez-vous mis en place des stratégies pour tout de même suivre les visiteurs qui ont refusé le tracking ?

Aucune à ma connaissance.

14. Au-delà des mises à jour techniques de vos outils (CMP, TMS) pour mieux collecter le consentement utilisateur sur les cookies/traceurs, avez-vous lancé un projet de réalignement de la gouvernance data au sein de votre entreprise ?

Pas à ma connaissance.

15. Quels impacts a eu le RGPD sur les données que vous utilisez pour piloter votre stratégie marketing ?

Il nous a fallu renforcer l'anonymisation des données collectées et utilisées dans l'ensemble de nos process et fournir un effort principalement technique (configuration outils) afin de nous assurer que les différentes chaînes d'exploitation des données que nous utilisons dans nos différents cas d'usage ne présentent pas de faille potentielle sur la sécurité des données personnelles des prospects et clients utilisant nos sites web.

16. Quels sont les impacts sur vos investissements média ?

Jusqu'à présent, pas d'impact significatif sur nos investissements. A l'avenir néanmoins des choix stratégiques pourront être effectués selon la façon avec laquelle nos principaux partenaires de l'ad tech rendront leurs solutions moins cookie-dépendants.

17. Prévoyez-vous ou avez-vous anticipé un éventuel report de budget média vers du SEO ou des partenariats ?

Pas à l'heure actuelle.

18. Est-ce que la mise en conformité au RGPD a engendré un coût budgétaire pour votre site ?

Pas à ma connaissance.

19. Identifiez-vous d'autres impacts que ceux évoqués précédemment concernant la mise en conformité du RGPD ?

Pas actuellement.

20. Pensez-vous que cette évolution de la législation changera drastiquement votre manière de travailler ?

Non, le suivi de la donnée restera toujours une préoccupation des entreprises, l'enjeu est surtout de comprendre comment celle-ci sera collectée et suivie demain afin d'adapter les processus et méthodes d'analyse actuels.

Annexe 4 - Questionnaire - Impact du RGPD sur Entreprise D

1. Quel est l'intitulé de votre poste et vos missions au sein de votre entreprise ?

Web Analyste

2. Comment suivez-vous les mises à jour des recommandations émises par la CNIL et les DPAs des autres pays (dans le cadre où votre entreprise est internationale) sur le RGPD ?

Veille, communication interne, éditeurs de solutions utilisées

3. Avez-vous dû créer une équipe dédiée pour suivre la mise en conformité au RGPD ?

Non

4. Si une équipe au sein de votre entreprise existe, de quel département dépend-t-elle (Sécurité, DSI Marketing, Juridique, etc.) ?

N/A

5. La CNIL a mis à jour ses recommandations en septembre 2020. Avez-vous anticipé la mise en place des nouvelles règles de la CNIL pour la collecte du consentement sur les cookies ?

Non, pas dans les faits, mais les réflexions sont menées en amont.

6. Si oui, quels sont les dispositifs que vous avez mis en place ?

N/A

7. Si non, quelles actions allez-vous mettre en place dans les prochains mois concernant la collecte du consentement pour les cookies ?

Utilisation de la mesure hybride avec une exemption de notre outil analytics, pour les autres cookies, gestion classique avec l'adaptation du bandeau notamment.

8. Pensez-vous que les différents acteurs (analytique, média, etc.) ont réussi à anticiper ces évolutions ?

L'outil analytics utilisé a bien anticipé certaines évolutions, car il s'agit déjà d'une solution qui bénéficiait de l'exemption depuis des années et qui a travaillé pour adapter et améliorer sa proposition.

D'un point de vue plus global, l'analytics mais aussi les autres acteurs ne sont peut-être pas dans l'anticipation ni même dans une réflexion sur des méthodes de collectes différentes à mon sens.

9. Traitez-vous le sujet de la collecte du consentement sur les traceurs à la fois pour vos sites web et vos applications mobiles ?

Oui

10. Avez-vous lancé des actions d'A/B testing de votre bandeau cookie et/ou de votre CMP afin d'optimiser la collecte du consentement ?

Oui

11. Quels sont les impacts du RGPD au niveau technique ?

Nous utilisons une CMP pour la gestion et l'autonomie sur les sujets

12. Est-ce que le tracking server-side est une solution que vous avez envisagée ?

oui complètement

13. Avez-vous mis en place des stratégies pour tout de même suivre les visiteurs qui ont refusé le tracking ?

Non, dans un premier temps nous nous baserons sur notre outil analytics et son suivi en mode exempté.

14. Au-delà des mises à jour techniques de vos outils (CMP, TMS) pour mieux collecter le consentement utilisateur sur les cookies/traceurs, avez-vous lancé un projet de réaligement de la gouvernance data au sein de votre entreprise ?

Non

15. Quels impacts a eu le RGPD sur les données que vous utilisez pour piloter votre stratégie marketing ?

Perte au niveau des tags media

16. Quels sont les impacts sur vos investissements média ?

Cela est géré par un autre département

17. Prévoyez-vous ou avez-vous anticipé un éventuel report de budget média vers du SEO ou des partenariats ?

Idem

18. Est-ce que la mise en conformité au RGPD a engendré un coût budgétaire pour votre site ?

Oui, coût au niveau des solutions utilisées et coût au niveau de l'investissement humain

19. Identifiez-vous d'autres impacts que ceux évoqués précédemment concernant la mise en conformité du RGPD ?

N/A

20. Pensez-vous que cette évolution de la législation changera drastiquement votre manière de travailler ?

A terme oui, il va falloir être inventif, innover pour continuer à travailler. Il est nécessaire de faire évoluer le métier en corrélation avec ces évolutions.

Annexe 5 - Politique de confidentialité

Introduction

Dans le cadre de mon projet de master Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales (MIASHS) parcours Web Analyse, je suis amené à collecter et à traiter des informations dont certaines sont qualifiées de "données personnelles". Moi, Guillaume Sinnaeve, attache une grande importance au respect de la vie privée, et n'utilise que des données de manière responsable et confidentielle et dans une finalité précise.

Données personnelles

Sur le site web www.guillaume-sinnaeve.com, il y a 1 type de données susceptible d'être recueilli :

- Les données collectées automatiquement

Lors de vos visites, je recueille des informations de type « web analytics » relatives à votre navigation comme la durée de votre consultation, votre localisation, votre type et version de navigateur. La technologie utilisée est sans cookies et sans JavaScript.

Utilisation des données

Les données « web analytics » sont collectées de forme anonyme par l'outil que j'ai développé, et me permettent de mesurer l'audience de mon site web, les consultations et les éventuelles erreurs afin d'améliorer constamment l'expérience des visiteurs. Ces données sont utilisées par moi-même, responsable du traitement des données, et ne seront jamais cédées à un tiers ni utilisées à d'autres fins que celles détaillées ci-dessus.

Base légale

Les données personnelles sont collectées sans consentement de l'utilisateur dans le cadre du respect des recommandations de la CNIL dans le cas de l'exemption.

Durée de conservation

Les données seront sauvegardées durant une durée maximale de vingt-cinq mois.

Contact délégué à la protection des données

Guillaume Sinnaeve - [sinnaeve.guillaume\[at\]gmail.com](mailto:sinnaeve.guillaume[at]gmail.com)

Annexe 6 - Événement de suppressions des données de plus de 25 mois

Ci-dessus, la requête utilisée permettant la mise en place d'un contrôle quotidien pour supprimer les données enregistrées il y a plus de 25 mois dans la base de données et ainsi être conforme aux recommandations de la CNIL.

```
CREATE EVENT `DELETE_DATA_OVER_25_MONTHS` ON SCHEDULE EVERY 1 DAY ON  
COMPLETION NOT PRESERVE ENABLE DO DELETE FROM `data` WHERE `timestamp` <  
CURRENT_TIMESTAMP - INTERVAL 25 MONTH
```

Annexe 7 - Campagnes d'acquisition de données

Texte de la publication en français

Bonjour à tous !

Dans le cadre de mon master MIASHS parcours web analyse, j'ai créé mon site internet expérimental (et non commercial) sur lequel j'ai mis en place une solution analytics sans cookies, sans JavaScript et conforme au RGPD à des fins de mesure d'audience. Découvrez mon site [□https://guillaume-sinnaeve.com/](https://guillaume-sinnaeve.com/)

J'ai besoin de vous pour générer du trafic et des actions sur le site afin d'alimenter l'outil en données et ainsi éprouver son efficacité.

N'hésitez pas à partager et à me contacter pour que je vous en dise plus sur cette solution.

Merci d'avance pour votre aide précieuse !

Texte de la publication en anglais

Hello everyone!

As part of my MIASHS web analysis course, I created my experimental (and non-commercial) website on which I set up an analytics solution without cookies, without JavaScript and GDPR compliant for audience measurement purposes. Discover my site [□https://guillaume-sinnaeve.com/](https://guillaume-sinnaeve.com/)

I need you to generate traffic and actions on the site in order to feed the tool with data and test its effectiveness.

Please do not hesitate to share and contact me so that I can tell you more about this solution.

Thank you in advance for your precious help!

Publication d'un post sur LinkedIn le mercredi 25 novembre 2020 à 12h00.



Guillaume Sinnaeve

Tracking & Analytics Manager chez fifty-five
maintenant •

...

Bonjour à tous !

Dans le cadre de mon master MIASHS parcours web analyse, j'ai créé mon site internet expérimental (et non commercial) sur lequel j'ai mis en place une solution [#analytics](#) sans [#cookies](#) et sans JavaScript à des fins de mesure d'[#audience](#).

Découvrez mon site [👉 www.guillaume-sinnaeve.com](https://www.guillaume-sinnaeve.com)

J'ai besoin de vous pour générer du trafic et des actions sur le site afin d'alimenter l'outil en [#données](#) et ainsi éprouver son efficacité.

N'hésitez pas à y faire un tour et à me contacter pour que je vous en dise plus sur cette solution.


Merci d'avance pour votre aide précieuse !

Guillaume Sinnaeve


guillaume-sinnaeve.com • Lecture de 1 min

It was in 2015 that I first encountered the world of data. And from the beginning, we liked each...

Publication d'un post sur Facebook le jeudi 26 novembre 2020 à 13h00.

**Guillaume Sinnaeve**
À l'instant · 🌐

Bonjour à tous !
Dans le cadre de mon master MIASHS parcours web analyse, j'ai créé mon site internet expérimental (et non commercial) sur lequel j'ai mis en place une solution [#analytics](#) sans [#cookies](#), sans [#JavaScript](#) et conforme au [#RGPD](#) à des fins de mesure d'audience.
Découvrez mon site 👉 <https://guillaume-sinnaeve.com/>
J'ai besoin de vous pour générer du trafic et des actions sur le site afin d'alimenter l'outil en données et ainsi éprouver son efficacité.
N'hésitez pas à partager et à me contacter pour que je vous en dise plus sur cette solution.
Merci d'avance pour votre aide précieuse !




GUILLAUME-SINNAEVE.COM
TMS and Analytics expert on the Web and Apps | Guillaume Sinnaeve
Guillaume Sinnaeve is an expert in data collection, web and applications, with and without TMS.

👍 J'aime

💬 Commenter

➦ Partager

 Votre commentaire...

😊 📷 GIF 🗨️

Publication d'un post sur LinkedIn le mardi 01 décembre 2020 à 14h30.



Guillaume Sinnaeve

Tracking & Analytics Manager chez fifty-five
maintenant • Modifié •



Hello everyone! I still need your help!

As part of my MIASHS web analysis course, I created my experimental (and non-commercial) website on which I set up an [#analytics](#) solution without [#cookies](#), without [#JavaScript](#) and [#GDPR](#) compliant for audience measurement purposes. Discover my site [👉 https://lnkd.in/dyK27qv](https://lnkd.in/dyK27qv)

I need you to generate traffic and actions on the site in order to feed the tool with data and test its effectiveness.

Please do not hesitate to share and contact me so that I can tell you more about this solution.

Thank you in advance for your precious help!

Guillaume Sinnaeve

guillaume-sinnaeve.com • Lecture de 1 min

It was in 2015 that I first encountered the world of data. And from the beginning, we liked each...



J'aime



Commenter



Partager



Envoyer

Publication d'un post sur Facebook le samedi 05 décembre 2020 à 13h00.



Guillaume Sinnaeve
À l'instant · 🌐

...

Bonjour à tous ! J'ai une nouvelle fois besoin de votre aide !

Dans le cadre de mon master MIASHS parcours web analyse, j'ai créé mon site internet expérimental (et non commercial) sur lequel j'ai mis en place une solution [#analytics](#) sans [#cookies](#), sans [#JavaScript](#) et conforme au [#RGPD](#) à des fins de mesure d'audience. Découvrez mon site 📄 <https://guillaume-sinnaeve.com/>

J'ai besoin de vous pour générer du trafic et des actions sur le site afin d'alimenter l'outil en données et ainsi éprouver son efficacité.

N'hésitez pas à partager et à me contacter pour que je vous en dise plus sur cette solution.

Merci d'avance pour votre aide précieuse !



GUILLAUME-SINNAEVE.COM

TMS and Analytics expert on the Web and Apps | Guillaume Sinnaeve

Guillaume Sinnaeve is an expert in data collection, web and applications, with and without TMS.

i

👍 J'aime

💬 Commenter

➦ Partager



Écrivez un commentaire...

😊

📷

🧠

🗨️

Annexe 8 - Schéma de la base de données

Schéma de la base de données

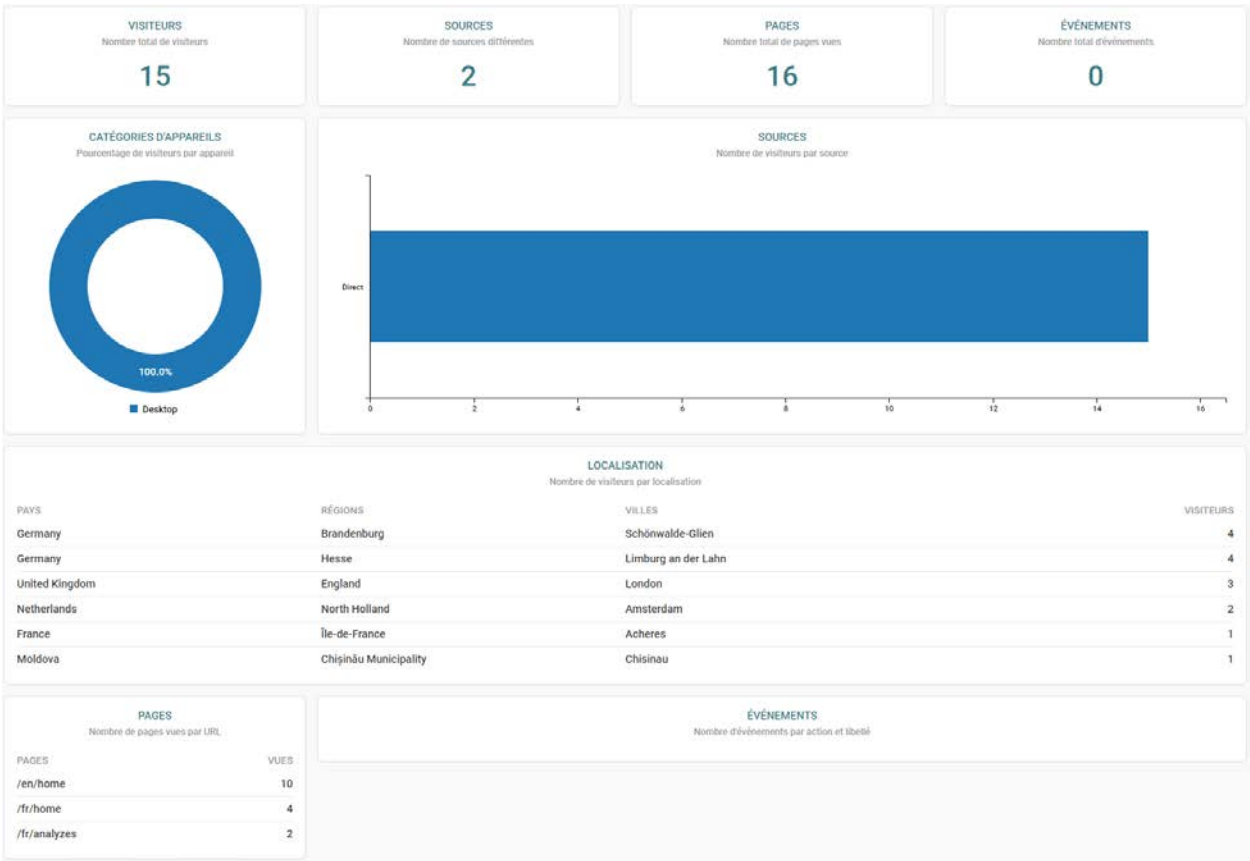
#	Nom	Type	Interclassement	Attributs	Null	Valeur par défaut
1	timestamp	timestamp			Non	CURRENT_TIMESTAMP
2	session_id	varchar(21)	utf8_general_ci		Oui	NULL
3	event_name	varchar(40)	utf8_general_ci		Oui	NULL
4	event_custom_1	varchar(100)	utf8_general_ci		Oui	NULL
5	event_custom_2	varchar(100)	utf8_general_ci		Oui	NULL
6	page_uri	varchar(512)	utf8_general_ci		Oui	NULL
7	date_full	varchar(10)	utf8_general_ci		Oui	NULL
8	date_day	varchar(16)	utf8_general_ci		Oui	NULL
9	date_hour	int(2)			Oui	NULL
10	date_minute	int(2)			Oui	NULL
11	referrer_full	varchar(512)	utf8_general_ci		Oui	NULL
12	referrer_host	varchar(128)	utf8_general_ci		Oui	NULL
13	referrer_path	varchar(128)	utf8_general_ci		Oui	NULL
14	geo_country	varchar(128)	utf8_general_ci		Oui	NULL
15	geo_region	varchar(128)	utf8_general_ci		Oui	NULL
16	geo_city	varchar(128)	utf8_general_ci		Oui	NULL
17	device_userAgent	varchar(256)	utf8_general_ci		Oui	NULL
18	device_category	varchar(32)	utf8_general_ci		Oui	NULL
19	device_system	varchar(32)	utf8_general_ci		Oui	NULL
20	device_browser	varchar(32)	utf8_general_ci		Oui	NULL
21	device_language	varchar(32)	utf8_general_ci		Oui	NULL

Liste des variables et des valeurs attendues :

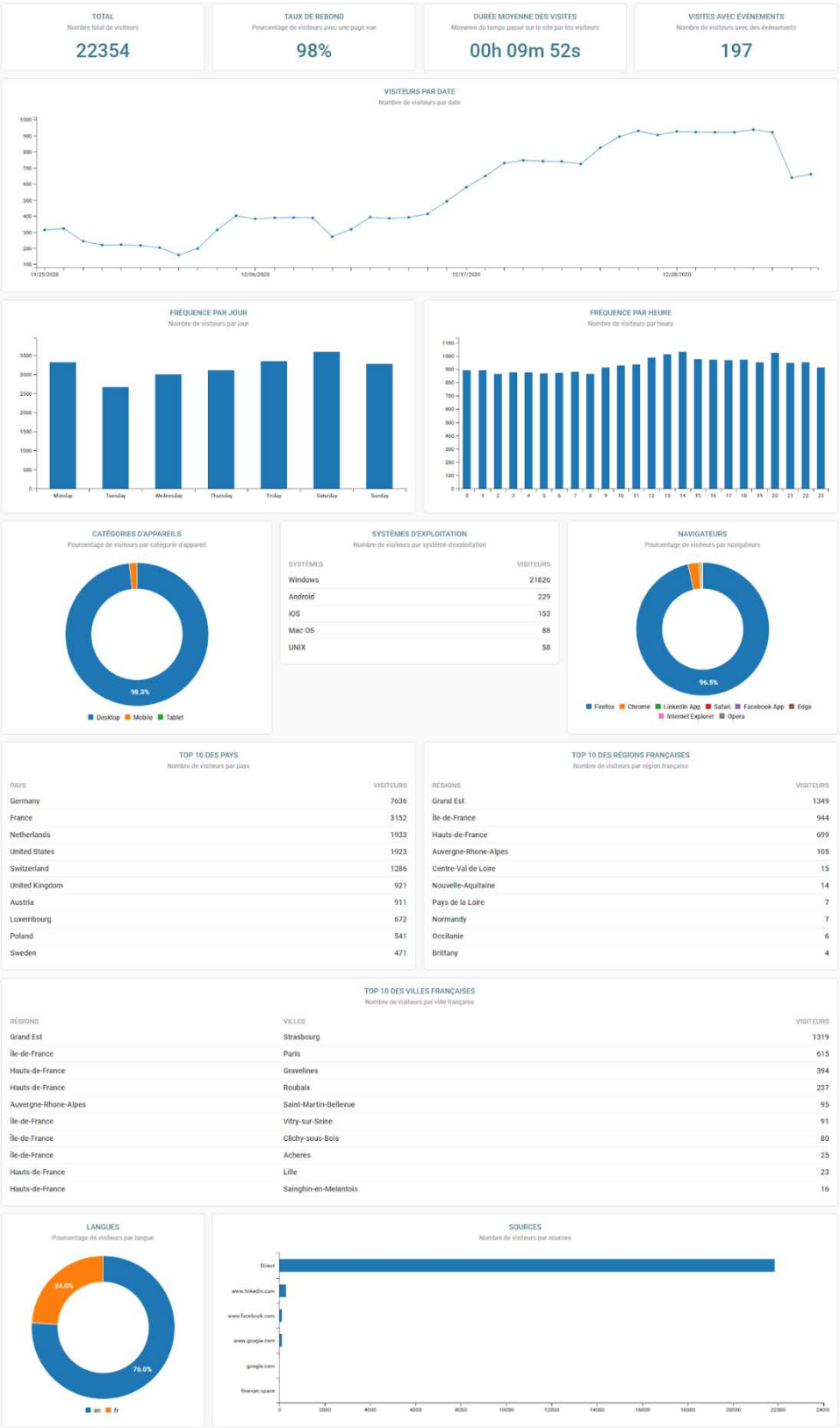
Variable	Description	Exemple
timestamp	Date et heure de la requête	2020-12-25 12:29:31
session_id	Identifiant de visite	1609170082_123456789
event_name	Nom de l'événement	show_ formations
event_custom_1	Qualificatif de l'événement	MIASHS
event_custom_2	Qualificatif de l'événement	University of Lille
page_uri	Chemin de la page	/en/formations
date_full	Date de la requête	12/25/2020
date_day	Jour de la requête	Friday
date_hour	Heure de la requête	12
date_minute	Minute de la requête	29
referrer_full	URL de la page précédente	https://www.guillaume-sinnaeve.com/en/home?sid=1609170082_123456789
referrer_host	Nom de domaine de la page précédente	www.guillaume-sinnaeve.com
referrer_path	Chemin de la page précédente	/en/home
geo_country	Pays du visiteur	France
geo_region	Région du visiteur	Île-de-France
geo_city	Ville du visiteur	Paris
device_userAgent	Agent utilisateur du visiteur	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/87.0.4280.101 Safari/537.36
device_category	Catégorie de l'appareil du visiteur	Desktop
device_system	Système d'exploitation du visiteur	Windows
device_browser	Navigateur du visiteur	Chrome
device_language	Langue du navigateur du visiteur	en

Annexe 9 - Page d'analyse du site

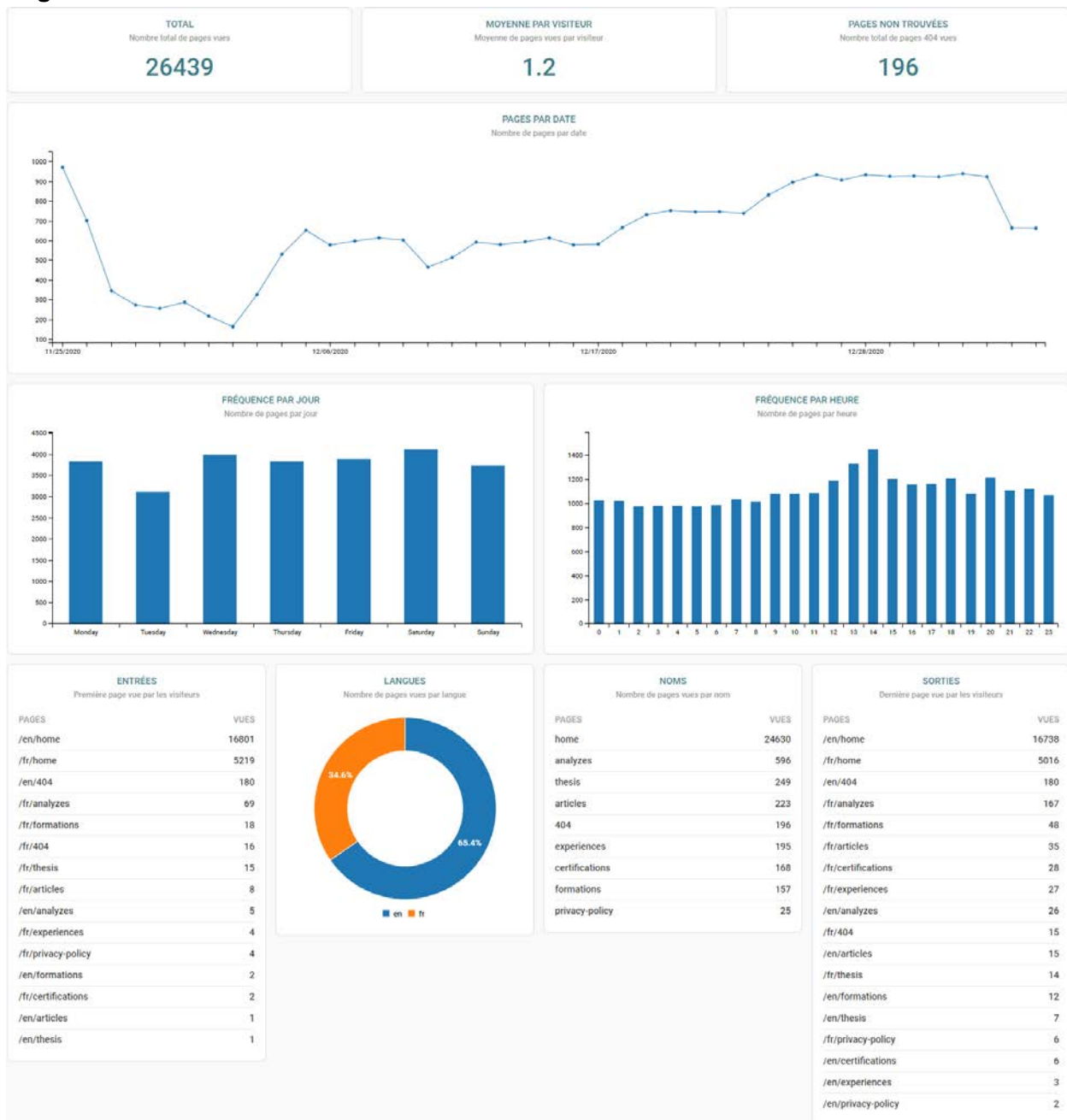
Temps réel



Visiteurs



Pages



PAGES PAR DATE
Nombre de pages par date

FRÉQUENCE PAR JOUR
Nombre de pages par jour

FRÉQUENCE PAR HEURE
Nombre de pages par heure

ENTRÉES
Première page vue par les visiteurs

PAGES	VUES
/en/home	16801
/fr/home	5219
/en/404	180
/fr/analyzes	69
/fr/formations	18
/fr/404	16
/fr/thesis	15
/fr/articles	8
/en/analyzes	5
/fr/experiences	4
/fr/privacy-policy	4
/en/formations	2
/fr/certifications	2
/en/articles	1
/en/thesis	1

LANGUES
Nombre de pages vues par langue

■ en ■ fr

NOMS
Nombre de pages vues par nom

PAGES	VUES
home	24630
analyzes	596
thesis	249
articles	223
404	196
experiences	195
certifications	168
formations	157
privacy-policy	25

SORTIES
Dernière page vue par les visiteurs

PAGES	VUES
/en/home	16738
/fr/home	5016
/en/404	180
/fr/analyzes	167
/fr/formations	48
/fr/articles	35
/fr/certifications	28
/fr/experiences	27
/en/analyzes	26
/fr/404	15
/en/articles	15
/fr/thesis	14
/en/formations	12
/en/thesis	7
/fr/privacy-policy	6
/en/certifications	6
/en/experiences	3
/en/privacy-policy	2

Événements



Table des illustrations

Illustration 1 : Schéma itératif de la collecte de données

Illustration 2 : Dépôt de cookies pour un visiteur sur un site e-commerce

Illustration 3 : Dépôt de cookies tiers suite à la vue d'une bannière publicitaire

Illustration 4 : Tendances de croissance trimestrielles des revenus publicitaires sur Internet 1996-2019. Source: IAB/PwC Internet Ad Revenue Report, FY 2019

Illustration 5 : Impact d'ITP sur les cookies first et les cookies tiers pour l'analytique et le média

Illustration 6 : Panneau de contrôle des différents niveaux de protection proposé par Firefox

Illustration 7 : Affichage sur le site www.urssaf.fr d'une pop-up de respect de l'option DNT activée sur le navigateur Google Chrome

Illustration 8 : Affichage du nombre de publicités et d'annonces bloquées sur le navigateur Brave sur un smartphone après 2 ans d'activité

Illustration 9 : Affichage d'une popin pour demander la désactivation des adblockers sur le site www.lesechos.fr

Illustration 10 : Schéma de la base de données collectées dans le cadre de l'expérience

Illustration 11 : Mention de la finalité des données collectées au sein de la page de politique de confidentialité

Illustration 12 : Création de l'événement de suppression automatique des données enregistrées au delà de 25 mois dans la base

Illustration 13 : Exemple d'une insertion du fichier collect.php sur une des pages du site

Illustration 14 : Visibilité de la ressource home.php dans le Network sur la page <https://www.guillaume-sinnaeve.com/fr/home>

Illustration 15 : Exemple d'en-têtes de réponse de la page <https://www.guillaume-sinnaeve.com/fr/home>

Illustration 16 : Styles CSS appliqués aux boutons de la page <https://www.guillaume-sinnaeve.com/fr/home>

Illustration 17 : Calculs effectués pour attribuer une valeur au paramètre “sid” de l’URL

Illustration 18 : Récupération de l’adresse IP pour déterminer le pays, la région et la ville de l’utilisateur

Illustration 19 : Récupération du nom de l’événement et des paramètres qui le qualifie via l’URL d’appel au fichier collect.php

Illustration 20 : Condition pour exclure les bots de la collecte de données

Illustration 21 : Calculs effectués pour déterminer les valeurs liées aux referrers

Illustration 22 : Calculs effectués pour déterminer le système d’exploitation de l’utilisateur

Illustration 23 : Rapports liés aux systèmes d’exploitations et aux navigateurs des visiteurs

Illustration 24 : Rapports liés aux langues et aux noms des pages vues

Illustration 25 : Rapport lié aux compétences consultés par les visiteur

Webographie

Aide Google Chrome. **Activer ou désactiver la fonctionnalité Interdire le suivi**

Adresse URL : <https://support.google.com/chrome/answer/2790761>

Benjamin Poilvé (Janvier 2020). **Une petite histoire du cookie**, Laboratoire d'Innovation Numérique de la CNIL

Adresse URL : <https://linc.cnil.fr/fr/une-petite-histoire-du-cookie>

CNIL. **Cookie**

Adresse URL : <https://www.cnil.fr/fr/definition/cookie>

CNIL (octobre 2020). **Cookies et traceurs : que dit la loi ?**

Adresse URL : <https://www.cnil.fr/fr/cookies-et-traceurs-que-dit-la-loi>

CNIL. **Définir une finalité**

Adresse URL : <https://www.cnil.fr/fr/definir-une-finalite>

CNIL (octobre 2020). **Mesurer la fréquentation de vos sites web et de vos applications**

Adresse URL : <https://www.cnil.fr/fr/mesurer-la-frequentation-de-vos-sites-web-et-de-vos-applications>

CNIL. **Sanction.**

Adresse URL : <https://www.cnil.fr/fr/definition/sanction>

Eur-Lex. **Directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (directive vie privée et communications électroniques).**

Adresse URL : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32002L0058>

Google. **Consent mode (beta)**

Adresse URL : <https://support.google.com/analytics/answer/9976101?hl=en>

IAB (Mai 2020). **Internet advertising revenue report Full year 2019 results & Q1 2020 revenues**

Adresse URL : https://www.iab.com/wp-content/uploads/2020/05/FY19-IAB-Internet-Ad-Revenue-Report_Final.pdf

ICO. **How do we comply with the cookie rules?**

Adresse URL : <https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/how-do-we-comply-with-the-cookie-rules/>

Journal du Net (Février 2020). **Privacy Sandbox : que contient l'alternative de Google aux cookies tiers ?**

Adresse URL : <https://www.journaldunet.com/ebusiness/publicite/1489199-privacy-sandbox-que-contient-l-alternative-de-google-aux-cookies-tiers/>

La Tribune (Novembre 2020). **RGPD : la CNIL condamne Carrefour à plus de 3 millions d'euros d'amendes**

Adresse URL : <https://www.latribune.fr/technos-medias/internet/rgpd-la-cnil-condamne-carrefour-a-plus-de-3-millions-d-euros-d-amendes-863384.html>

Le Monde (Avril 2018). **Données privées : le site de rencontres Grindr mis en cause**

Adresse URL : https://www.lemonde.fr/pixels/article/2018/04/03/donnees-privees-le-site-de-rencontres-grindr-mis-en-cause_5279794_4408996.html

Les Echos (Janvier 2018). **Criteo : un pépin nommé Apple**

Adresse URL : <https://www.lesechos.fr/2018/01/criteo-un-pepin-nomme-apple-981952>

Radio Canada (Juin 2019). **Fuites de données : cinq grands scandales des dernières années**

Adresse URL : <https://ici.radio-canada.ca/nouvelle/1193991/scandale-fuite-vol-renseignements-personnel>

RGPD. **CHAPITRE I - Dispositions générales**

Adresse URL : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre1>

Romain Warlop (Février 2020). **Privacy Sandbox : quand Google relève le défi de l'alternative aux cookies publicitaires... en deux ans !**, 55 the tea house

Adresse URL : <https://teahouse.fifty-five.com/fr/privacy-sandbox-quand-google-releve-le-defi-de-l-alternative-aux-cookies-publicitaires-en-deux-ans/>

Stratégies (Novembre 2018). **Les adblockers ne connaissent pas la crise**

Adresse URL : <https://www.strategies.fr/actualites/medias/4020718W/les-adblockers-ne-connaissent-pas-la-crise.html>

The Mozilla Blog (Août 2020). **Latest Firefox rolls out Enhanced Tracking Protection 2.0; blocking redirect trackers by default**

Adresse URL : <https://blog.mozilla.org/blog/2020/08/04/latest-firefox-rolls-out-enhanced-tracking-protection-2-0-blocking-redirect-trackers-by-default/>

Webkit (Juin 2017). **Intelligent Tracking Prevention**

Adresse URL : <https://webkit.org/blog/7675/intelligent-tracking-prevention/>

Webkit (Juin 2018). **Intelligent Tracking Prevention 2.0**

Adresse URL : <https://webkit.org/blog/8311/intelligent-tracking-prevention-2-0/>

WebKit (Février 2019). **Intelligent Tracking Prevention 2.1**

Adresse URL : <https://webkit.org/blog/8613/intelligent-tracking-prevention-2-1/>

WebKit (Avril 2019). **Intelligent Tracking Prevention 2.2**

Adresse URL : <https://webkit.org/blog/8828/intelligent-tracking-prevention-2-2/>

WebKit (Septembre 2019). **Intelligent Tracking Prevention 2.3**

Adresse URL : <https://webkit.org/blog/9521/intelligent-tracking-prevention-2-3/>