GOUVERNEMENT

Liberté
Égalité
Fraternité

Pôle d'expertise de la régulation numérique

The Privacy Sandbox: a collection of tools for third-party cookie-less online advertising

Today, the online advertising ecosystem heavily relies on the use of third-party cookies to improve the relevance of ads presented to users. Google has decided to remove these third-party cookies from its market-leading Chrome browser by 2023. It is proposing a technological alternative called " The Privacy Sandbox", a collection of tools presented by Google as both more privacy-friendly and economically viable.

The Privacy Sandbox is currently under active development and will continue to evolve. This issue of *Shedding light on…* provides an assessment on the current state of this initiative as of March 22. As of this date, not all use cases permitted by third-party cookies are reproducible by The Privacy Sandbox. It has not yet demonstrated its ability to enable companies to monetise advertising as well as they could with third-party cookies. In this context, Google could benefit from a substantial comparative advantage. Google, directly hosting more and more content, would be less affected than its competitors that rely more heavily on third-party cookie targeting.

| Shedding light on…

March 2022

#03

KEY CONCEPTS TO UNDERSTAND THE PRIVACY SANDBOX INITIATIVE

Online advertising: adapting the message to the context and the user

According to CNIL[1], targeted (or personalised) advertising[2] is "an advertising technique that aims at identifying people individually in order to deliver specific advertising messages based on individual characteristics". The Privacy Sandbox initiative addresses the broader framework of *programmatic advertising*. This includes the so-called contextual advertising, based on the content displayed on the web page, or behavioural targeting advertising, if the segmentation is based on a user's browsing history. Being able to target users (e.g. by identifying their interests) as they browse serves the entire online advertising ecosystem to buy, track, analyse, measure, control and adapt ad campaigns in real time.

Programmatic advertising
Allows to plan the purchase of advertising space automatically according to predefined criteria (price, audience characteristics, time of day, etc.).

The global online advertising market was worth USD 378 billion[3] in 2020 and largely finances the so-called free web (applications; websites, for example, certain blogs and journalistic articles). Targeted advertising accounts for a significant part of the generated revenues, and it relies massively on collecting and using data from users.

Google earns the majority of its revenue from online advertising[4]. It is a major player in this field, if not the dominant one. Recently, Google said that it now has to reconcile protecting its users' privacy from third parties with the economic impact that the removal of third-party cookies could have on the ecosystem. The Privacy Sandbox initiative therefore aims at enabling targeted programmatic advertising to continue operating in a world without third-party cookies, while preserving the privacy of users from third-party trackers.

Cookies: a core technology for online advertising

To target users today, **the online advertising ecosystem relies on the massive use of cookies**. These text files are saved in the web browser and are associated with a domain name[5], making it possible to store data about users' navigation and actions. Two types of cookies are used: first-party cookies and third-party cookies.

First-Party
By first-party we mean an actor who interacts directly with their audience. Thus, the data collected by Cdiscount as a user navigates on cdiscount.com are said to be first-party.

First-party cookies are set by the visited website, and are mainly intended to enable the site to properly function and to improve the user's experience (for example, to remain logged in, to maintain a marketplace

[1] French Data Protection Authority (https://www.cnil.fr/en/home)

[2] CNIL, Targeted advertising definition (https://www.cnil.fr/fr/definition/publicite-ciblee)

[3] Statista, Digital advertising spending worldwide from 2019 to 2024 (in billion U.S. dollars), May 2021 (https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/).

[4] Approximately 80% of its revenues, $147B in 2020. Alphabet Inc., Alphabet Announces Fourth Quarter and Fiscal Year 2020 Results, February 2021 (https://abc.xyz/investor/static/pdf/2020Q4_alphabet_earnings_release.pdf).

[5] For example: lemonde.fr, more details available on Afnic.fr (https://www.afnic.fr/noms-de-domaine/tout-savoir/).

cart, to carry out audience measurements, etc.). In addition, the website may retain data about the users and their navigation, even if they did not log in.

In addition to the direct interaction between a website and a visitor, a web page may display content from other websites (e.g. videos, images, podcasts, fonts, etc.). These third-party websites are then able to place and read their own cookies on the user's browser, **which are known as "third-party" cookies**.

Example: if a user browses from site A to blog B, both of which display videos hosted on third-party website C, then site C will be able to place and read its own cookies on the user's browser. Server C will then know the user's browsing history wherever C was a third-party, in this case site A and then blog B.

This mechanism has led to the rise of companies specialised in tracking users. As some of them have managed to partner with other companies, their cookies are now embedded in most of the visited websites, enabling them to be aware of a major proportion of users' browsing history. These histories can then be analysed to deduce users' preferences and profiles, which can finally be used for targeted advertising.

Websites that were directly visited by users, such as A or B, agree to let sites such as C track their users, as it allows them to better monetise ads on their pages in exchange. **Thus, tracking users via third-party cookies allows companies to build profiles by using a wide range of data, including personal data. Even though this tracking requires users' consent under GDPR, it lacks transparency and remains difficult to control, as does the choice of what data can be transmitted for what purpose.**

Towards the end of third-party cookies?

On grounds of privacy considerations, Mozilla and Apple gradually restricted third-party cookies in their respective browsers: Firefox and Safari[6]. For example, in 2019, Firefox blocked third-party cookies identified as tracking-related, and in 2021 confined all third-party cookies to the site they were created on, preventing tracking companies to track users from site to site. Apple, on its side, initiated a restrictive blocking of third-party cookies in 2017 in Safari with the first iteration of webkit's Intelligent Tracking Prevention.

[6] Mozilla, Firefox now blocks tracking third-party cookies and cryptomining cookies by default, 3 September 2019 (https://blog.mozilla.org/press-fr/2019/09/03/firefox-bloque-desormais-par-defaut-les-cookies-tiers-de-pistage-et-les-mineurs-de-cryptomonnaies/). Sabharwal, C. Safari ITP: Intelligent Tracking Prevention Version 1.0 to 2.3. Adpushup, 10 September 2020 (https://www.adpushup.com/blog/safari-itp-intelligent-tracking-prevention/).

In both cases, restricting third-party cookies is estimated to have translated into revenue decreases for targeted advertising companies[7] as well as for websites themselves.

W3C
World Wide Web Consortium, non-profit standardization organization. Works on the development of web standards such as HTML, CSS…

Google, being the market leader with its Chrome Internet browser (70% market share in 2021[8]), launched in 2019 an open source initiative within the World Wide Web Consortium (W3C) standardisation forum **This initiative, called Privacy Sandbox[9], aims at implementing in Chrome by 2023 a set of advertising tools to replace third party cookies in a way that meets all the advertising use cases enabled by third-party cookies, while being better at protecting users' privacy (see below).**

API
Application Programming Interface a software interface that allows one software or service to be "connected" to another software or service in order to exchange data and functionality.

PRIVACY SANDBOX: A PROJECT CLOSELY MONITORED BY ALL STAKEHOLDERS

In the wake of growing privacy considerations and various scandals[10], Mozilla and Apple have limited access to their users' data, at the expense of the advertising industry's and indirectly publishers' revenues[11].

In turn, Google started working on The Privacy Sandbox in 2019. While the original release date was supposed to be in 2022, Google pushed it back to at least 2023. Each tool in The Privacy Sandbox project is publicly described, and a list of the tools is available on the project's website[12]. The Privacy Sandbox initiative includes several categories, such as: "Show relevant content and ads", "Measure digital ads" and "Fight spam and fraud on the web"[13]. The discussions around these tools are hosted in W3C working groups, of which Google is an influential member.

In the first half of 2020, the tools went rapidly through the incubation phase (see below for their description), and some of them entered the experimentation phase in 2021. However, these experiments led the community of adtech to express doubts[14] on the consequences of the technologies and their effects on competition. In June 2021, the UK's Competition and Markets Authority (CMA)[15], fearing that The Privacy

[7] Hern, A. No tracking, no revenue: Apple's privacy feature costs ad companies millions. The Guardian, 9 January 2018 (https://www.theguardian.com/technology/2018/jan/09/apple-tracking-block-costs-advertising-companies-millions-dollars-criteo-web-browser-safari).

[8] CNIL, 13 October 2021 (https://www.cnil.fr/fr/alternatives-aux-cookies-tiers-quelles-consequences-en-matiere-de-consentement)

[9] Privacy Sandbox website (https://privacysandbox.com/)

[10] Ad Industry Accused Of 'Massive' Privacy Breach, Forbes, 18 January 2019 (https://www.forbes.com/sites/emmawoollacott/2019/01/28/ad-industry-accused-of-massive-privacy-breach/).

[11] Ibid, 7.

Note: On that matter, complaints have been filed against Apple, alleging the company would have an ulterior motive of a advantaging its own advertising services instead of wanting to protect its users' privacy. These complaints are being investigated by the French Antitrust Authority (ADLC)

[12] Ibid, 9.

[13] Privacy Sandbox Overview, as of 31 March 2022 (https://privacysandbox.com/open-web/#how-works-on-web-hero)

[14] WIRED, Antitrust and Privacy are on a collision course, 12 April 2021 (https://www.wired.com/story/antitrust-privacy-on-collision-course)

[15] Competition and Market Authority (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/992975/Notice_of_intention_to_accept_binding_commitments_offered_by_Google_publication.pdf).

Sandbox would strengthen Google's position in the advertising ecosystem[16], initiated a formal procedure to review the initiative. Google reacted by offering binding commitments to address the CMA's concerns and promised better transparency around its initiative. An official provisional timeline for the project was published[17]. In November 2021, after a consultation process conducted by the CMA with various advertising companies, Google extended the scope of its commitments, though without fully addressing the concerns raised by the British authority[18].

At the same time, Criteo, a French company specialised in targeted advertising, questioned The Privacy Sandbox's ability to enable online advertising companies to continue their activities with the same functionality as those allowed by third-party cookies[19]. Criteo indeed conducted tests on FLoC (Federated Learning of Cohorts)[20] - a tool that was part of The Privacy Sandbox at the time, and concluded notably that The Privacy Sandbox would probably not yield as much revenue for stakeholders as current technologies do.

ONLINE ADVERTISING VALUE IMPLICATIONS

The Privacy Sandbox is not yet finalised and could still be modified before its release in order to correct the identified limitations, some of which are listed below.

Not all use cases enabled by cookie technologies are yet reproducible with The Privacy Sandbox tools. For example, it is not possible to use video ads[21] via FLEDGE[22], an advertising targeting tool.

Monetization
Free online content (press, videos…) is often financed through online advertising. We talk about content monetization when the editor gets paid without making the user pay but by offering him ads.

The ability of Privacy Sandbox solutions to monetise as well as existing cookie technologies is not proven. Indeed, if there is less individual tracking and less functionality available than with third-party cookies, there will probably be less opportunities to create value. Furthermore, if Privacy Sandbox fails to monetise as well as third-party cookie technologies do, this would only have consequences for players that depend on third-party cookie technologies. Players relying on first-party cookies, particularly in environments requiring user authentication, will not be affected by this technological change, as the change does not hinder the ability of companies to collect data via first-party cookies. Companies that can work without third-party cookies, typically via a logged-in environment, will not be harmed in any way and will even experience less competition from the third party cookie players.

[16] For example, see article 5.3 from the CMA's notice (Ibid, 15).

[17] Privacy Sandbox timeline (https://privacysandbox.com/open-web/#the-privacy-sandbox-timeline)

[18] See article 2.3 and 2.4 of CMA's analysis and Google's new commitments (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1036204/211126_FINAL_modification_notice.pdf).

[19] « *FLoC Origin Trial* » series on Criteo Tech's blog, Medium , link to the first article (https://medium.com/criteo-engineering/floc-origin-trial-what-we-observed-3f7e8f209b82).

[20] *FLoC* on GitHub (https://github.com/WICG/floc).

[21] Video advertising on the web #29, GitHub (https://github.com/WICG/turtledove/issues/29).

[22] FLEDGE on GitHub (https://github.com/WICG/turtledove/blob/main/FLEDGE.md)

In this context, a substantial advantage for Google seems to be emerging: being increasingly in a first-party position, Google would be less affected than its competitors, should The Privacy Sandbox turn out to be less effective than existing cookie technologies. Indeed, Google is increasingly hosting content directly on its own servers, as it pushes to provide ever-faster response times to its users. Google does so, for example, by integrating responses directly into its results pages (knowledge cards, weather answers, theatre show times, etc.) or by promoting content in AMP (Accelerated Mobile Pages) format, that enables the search engine to preload and cache publishers' content on its servers. Hosting this content allows Google to track all users' activities as a first party, without losing any information. Conversely, the owner of the content distributed via AMP, becomes a third party and depends on Google to access the most detailed information about the users reading their own content.

Finally, even if there were no costs for the industry to transition and adopt the Privacy Sandbox and if it generated the same revenues as current technologies, there would still be an economic risk for websites and targeted advertising companies that could stem from **a change in the distribution of value between companies.** As the mechanisms of The Privacy Sandbox are designed to prevent collection and cross-referencing of data between websites, some companies such as Data Management Platforms or DMPs[23] will probably see a decrease in their activity.

PRESENTATION OF TECHNICAL COMPONENTS OF THE PRIVACY SANDBOX

In order to make the changes brought by The Privacy Sandbox more concrete, this section presents four of the project's tools:

- FLoC, Topics and FLEDGE which address the need to present relevant advertisements to the user;
- Trust Token API which is a technical solution to fight fraud.

Other tools included in The Privacy Sandbox address different areas such as campaign reporting and the fight against cross-site tracking.

Tracking
The practice of associating a unique identifier with the user in order to track their navigation through the sites.

Interest-based targeting solutions: from FLoC to Topics

Interest Based Advertising (IBA) is based on the collection of data on web domains. Ads presented to the user are tailored to their known preferences or interests which can be inferred from their browsing history.

Google presented a first iteration of a targeting tool: FLoC[24]. The proposed approach consisted in grouping users into cohorts according to their browsing history similarity. Each cohort corresponds to a mix of interests deduced from the visited domain names. To ensure privacy, the

[23] CNIL, Data management platform (DMP) (https://www.cnil.fr/fr/definition/data-management-platform-dmp-ou-plateforme-de-gestion-des-donnees)

[24] Ibid. 20.

decision to assign users to a cohort was to be periodically and locally made by the browser.

From a privacy perspective, the choice of relying on a cohort number has been heavily criticised,[25] as it could be used as an entry point to identify users with fingerprinting mechanisms. Moreover, since processing this identifier requires the user's consent under GDPR, Google was unable to conduct its test phase in countries where the GDPR applies, hindering the stakeholders' ability to analyse this solution.

Fingerprinting
A technique to uniquely identify a user based on information provided by their hardware. For example, their IP address, the version of their browser…

For online advertising players, the cohort number did not directly provide usable semantic information and therefore required more work and investment to translate it into human-readable interests. This could have resulted in an increased asymmetry between big and small players, the latter being less able to make this translation. **At the same time, as any company could access the cohort number, this system tended to spread this piece of information among all firms for free**: content publishers could not monetize with whom they wanted to share the information that a user visited their website.

Following these numerous criticisms, a new tool was proposed in January 2022 to replace FloC: Topics[26]. The new mechanism relies on the browser associating the users with their 5 favourite topics every week, based on their browsing history. These top 5 topics are kept for 3 weeks by the browser, which can then record a maximum of 15 topics.

As of today, the complete list of interests is predetermined in order to exclude sensitive categories. The list will evolve in the future and currently includes 349[27] topics. A machine learning model residing locally in the browser analyses the names of recently visited domains to determine the prevalent topics associated with visited websites. As each site is associated with one or more topics, the user's browser counts the number of times those topics have been encountered at each website visit. At the end of each week, the five most encountered topics constitute the "favourite" interests of a user.

Ad companies only have access to a subset of the list of 349 topics, (i.e., the ones associated with their own or their partners' websites), which constitutes their own observable list of interests. They only can receive information about a (anonymous) user having a particular interest if this user has previously visited their own or their partner's websites, and if the topic sent back by the Topics API is part of their own observable topics list. This mechanism mimics that of third-party cookie technologies in that partnerships between stakeholders and websites (or owning multiple websites) are necessary to gain access to information.

[25] Electronic Frontier Foundation, Google FLoC Is a Terrible Idea, March 3, 2021 (https://www.eff.org/fr/deeplinks/2021/03/googles-floc-terrible-idea).

[26] Topics on GitHub (https://github.com/jkarlin/topics).

[27] Topics taxonomy on GitHub (https://github.com/jkarlin/topics/blob/main/taxonomy_v1.md).

In practice, when users visit a website, the topics engine in their browser will randomly choose three (3) topics from the 15 registered ones. Each company present on the visited website (the site owner and their partners) will then be able to receive these interests during one week, if they previously observed the user and if the topics randomly selected are present in their observable list.

The concept of interests at the core of Topics is way more usable than the hardly decipherable cohort number that was provided by FLoC. Nevertheless, the "top 5" mechanism, combined with uncertainties about the generalization or extension of the topics list, casts doubts about the ability to have topics that are narrow enough for effective targeting. Consequently, ad personalisation and the added value generated for the whole ecosystem might be restrained. Contrary to what was planned for FLoC, Topics computation will only be activated on sites that actively use the tool, but it will be activated by default on the user side.

Retargeting
A marketing strategy that allows an advertiser to specifically target visitors who have not yet taken the desired action (typically a purchase).

Advertising (re)targeting solution: FLEDGE

FLEDGE[28] aims at enabling advertisers to display ads to "potentially interested" users who have already interacted with the advertiser's or one of its partner's websites. More precisely, FLEDGE offers, on the one hand, a solution for creating and updating **interest groups** and, on the other hand, **a bidding mechanism** residing in the user's browser. It is critical to distinguish FLEDGE "interest groups" from Topics' "interests", as FLEDGE's interests groups are merely groups of users sharing a similar action on a website or a group of websites (for example all users who visited a same website or who put a specific item in a shopping cart).

During its consultative approach, Google used the feedback from many stakeholders[29] on its initial tool, TURTLEDOVE, to develop its replacement one, FLEDGE.

FLEDGE is built so that the browser operates the auctions and stores information about the user's membership to interest groups. The website on which the advertisement is to be displayed should therefore not be able to know which groups participated in the auction nor their biddings. Advertisers for their part can only base their bids on the interest group data and are theoretically prevented from identifying any individual user in any group. Otherwise, advertisers would be able to cross-reference the user's interests with other information.

FLEDGE is at the core of The Privacy Sandbox because it addresses common use cases in online advertising. Some examples are:

- An online shop can easily carry out advertising retargeting. To do so, it adds its visitors to an interest group.
- A car forum can qualify its audience as being interested in cars. It allows third parties to target its car enthusiast interest group.

[28] FLEDGE on GitHub (https://github.com/WICG/turtledove/blob/main/FLEDGE.md#summary).

[29] RTB House, NextRoll, Magnite, Criteo, and Google Ads team.

- A publisher can delegate the creation of interest groups to another company specialised in user qualification. At first sight, this mechanism strengthens the position of the publisher sites by keeping them in control of their data.

Nevertheless, a company could partner with several publishers to create and add users to a single group on several websites at once, giving it a more favourable position to create large groups. It is possible that in the end, the advantage will go to the intermediaries — for example the Sell Side Platforms — giving them a more favourable position to create and populate groups.

FLEDGE is still under development. Its challenge is to be able to carry out an advertising auction within the browser while limiting the access of companies to user data. In order to perform the auction, the browser must exchange certain information with the advertising players[30]. Without an appropriate mechanism, these exchanges could be leveraged by players to collect or deduce information about users. To prevent that, FLEDGE will use a decentralised[31] server mechanism that still needs to be defined. In the meantime, the first experimentation will use the servers of the involved[32] stakeholder, which considerably limits the guarantees sought in terms of privacy protection. Without further information, it is difficult to predict how this technical challenge will be met.

Transferring information about the authenticity of an Internet user: Trust Token API

Trust Token API is a tool proposed by Google to share the trust that a site establishes in a user, such as determining their authenticity, for example to ensure that they are not a bot.

Bot
Automated computer program whose purpose is to simulate the behaviour of a human being and perform tasks. In advertising, bots can be used to generate fake traffic on a site, thus generating fake advertising costs for advertisers.

The theoretical advantage is twofold:

- For websites, to have guarantees of traffic authenticity at a lower cost.
- For users, to reduce the number of captchas or other means of checking their legitimacy.

This project, which has no direct implications in terms of targeted advertising, nevertheless risks reinforcing the ability of certain players, such as Google, to monitor the evolution of traffic on many websites and in real time.

In a simplified view, **this tool distinguishes between two roles: on the one hand, an issuer website, which verifies the legitimacy of users, and on the other hand, the redeemer website, which wants to check the user's trustworthiness**. The trust would be conveyed through an anonymous

[30] For example, the browser could query the advertiser, associated with an interest group, for the campaign remaining budget.

[31] The 'Trusted-Server' model includes a server that holds the information needed for the auction, but keeps no record of exchanges with users. It is not yet defined.

[32] "Bring Your Own Server": during the experimentation, each actor will be allowed to exchange data with their own server.

"token" provided to the user by an issuer. When the user visits another site - called a redeemer - it asks the issuer to validate the authenticity of the token (to avoid a build-up of such tokens or their fraudulent use).

Endorsing the role of issuer presents many technical complexities (to limit the risks of attacks and to improve the detection of bots), as well as potentially significant costs (development, servers). Redeemer sites have an interest in choosing tokens from issuers that are better at identifying legitimate users, which can lead to competition between them. However, no remuneration mechanism for companies endorsing the redeemer role seems to be considered. Thus, a Facebook engineer[33] expressed doubts about the interest of companies to take on the role of issuer, but also about the risks regarding users' privacy, for example a user presenting a token issued by a given platform would necessarily be a user of this same platform.

All these requirements mean that the role of token issuer may not be very attractive for most players. Paradoxically, it could be particularly interesting for others, including Google. Indeed, the issuance of tokens and the real-time feedback of trust requests could provide these issuers anonymous but very detailed information on the sites visited by users and their traffic. This knowledge could be an advantage in related markets (useful for ranking in search engines, strategic information on competitors' audiences, etc.). Furthermore, we cannot exclude that this service, for now presented as free, will eventually evolve into a paid service, either for the users or for the visited websites (which want to know that their visitors are legitimate). If it becomes impossible or too complex to guarantee the authenticity of a user without these trust tokens, issued by a probably limited number of actors, monetisation cannot be excluded in the medium or long run.

AN EXTENSION OF THE PROJECT: THE PRIVACY SANDBOX ON ANDROID

In February 2022, Google announced that it wanted to extend The Privacy Sandbox project to the mobile app ecosystem. Its objective is similar to its web project counterpart: providing applications that enable sustainable revenues for advertisers and publishers while limiting tracking and especially the use of the Google Advertising ID[34], a unique identifier used for advertising purposes.
Google is giving itself two years to adapt some of the tools of the web Privacy Sandbox and to develop new ones[35]. Participation in this initiative will take place directly in the developer section of the Android[36] website.
The mobile privacy sandbox will be a separate project and will require further analysis as it progresses.

[33] On GitHub (https://github.com/WICG/trust-token-api/issues/28)

[34] Google Play (https://support.google.com/googleplay/android-developer/answer/6048248)

[35] For instance, SDK Runtime (https://developer.android.com/design-for-safety/ads/sdk-runtime).

[36] Privacy Sandbox on Android (https://developer.android.com/design-for-safety/ads).

Legal deposit: March 2022
ISSN (online): 2824-8201
contact.peren@finances.gouv.fr

The Pôle d'Expertise de la Régulation Numérique (PEReN) is a national service providing expertise and technical assistance in the fields of data processing, data science and algorithmic processes to government departments and administrative authorities involved in the regulation of digital platforms. It is also involved in exploratory and scientific data science research projects

PEReN – 120 rue de Bercy, 75572 Paris Cedex 12