

# Appendix G: the role of tracking in digital advertising

## Summary

1. This appendix describes the tracking of users and their devices for personalised advertising and attribution, including:
  - (a) some of the main technologies currently in use for tracking;
  - (b) the limitations of current controls available to users and the technical challenges in limiting tracking;
  - (c) applications of tracking technologies in adtech, in particular by Google and Facebook;
  - (d) estimates of the prevalence and prominence of tracking on the internet and in mobile apps; and
  - (e) some recent and near-future developments affecting tracking, such as restrictions on third-party cookies and proposals to limit cross-site tracking using privacy-enhancing technologies.
2. It serves as a reference for setting out our understanding of how tracking works to support the conclusions in our main report.
3. Websites and apps may collect data on a user for many purposes, such as to measure and improve their service, or to detect fraud and abuse. In this appendix, we focus on cross-property tracking for personalised advertising purposes. We define tracking as the collection and linking of data on a user or device across multiple websites or applications, and the retention, use, or sharing of that data.<sup>1</sup>
4. The goal of tracking for personalised advertising is to link together the activity of a single user across different sessions, properties (webpages and apps) and devices, to build a more complete profile of that user. This user profile could, among other things, help deliver personalised advertising and inform advertisers' spending decisions. We note that personalised advertising currently relies on tracking methods not only to target individuals, but also for attributing conversions and evaluating the effectiveness of advertising.
5. Identifying users is crucial to tracking. There are a wide range of tracking methods, the most well-known of which are third-party cookies. This appendix

---

<sup>1</sup> This definition of tracking is adapted from [Disconnect](#). It differs from Disconnect's in that we include cases where users are tracked across multiple properties owned by the same company, because a user may not know which properties are owned by whom, and what their internal data re-use policies are.

explains core concepts about identifiers and identification, and a selection of current methods for tracking - covering web, mobile, and cross-device tracking. We discuss how these tracking methods exploit the fundamental way in which commonplace technologies, used by consumers in everyday life, work. These technologies include HTTP (the protocol of the WWW), the use of third-party libraries (TPLs) in websites and applications and the fact that pre-installed software in an open source operating system can rewrite permissions to user and device data.

6. Notwithstanding efforts to better inform users about tracking and to obtain their valid consent since GDPR, it is likely that most typical users are unaware of the full extent to which they may be tracked, and are not in a position to make informed decisions or to take actions (including technical measures) that limit tracking. In some cases, users face a choice to either accept tracking or to stop using many services and technologies altogether. We set out and explain some of the controls available to users at the browser and mobile OS level, and their limitations. Platform and provider-level controls are set out more fully in Appendix K.
7. Devices, operating systems and browsers can play critical roles influencing user behaviour and limiting tracking. Major browsers and mobile operating systems (OSs) have been developing technologies to restrict tracking and data sharing, including by default; Apple's Intelligent Tracking Prevention (ITP) is one example. Tracking methods and privacy-protecting technologies have and will continue to develop in an ongoing arms-race. In addition to platforms or browsers, we look at how tracking technologies have been influenced by internet standards settings organisations (such as the IETF, W3C and WHATWG) or by browsers working together.
8. This appendix explains several applications of tracking technologies and use of data in the adtech ecosystem, including potential data protection concerns. This includes the tracking inherent to real-time bidding in the broadcasting of bid requests and cookie (and other identifier) matching, but also supporting services such as integrations offered by data management platforms (DMPs) and data brokers, and the use of tracking technologies and data from tracking by Google and Facebook and their customers.
9. Currently, tracking is necessary for many activities which enhance the efficiency of personalised advertising, such as targeting and user-level attribution. We present an overview of estimates of the extent to which tracking technologies are found in websites and mobile apps. Not only do Google and Facebook operate leading consumer-facing services that provide rich data and insights for targeting about a large proportion of consumers; the evidence suggests that Google and Facebook also have the widest

prevalence of tracking technologies on websites and mobile apps, allowing them to track users across much of the internet. We posit a ‘feedback loop’ between tracking and market power. Large incumbent platforms, such as Google and Facebook, have greater opportunities to track users, which improves the effectiveness of their targeting and personalised advertising. This encourages publishers and advertisers to use Google’s and Facebook’s adtech services, which allows Google and Facebook to track users on more websites and apps, giving these platforms yet more presence and opportunities to track users across the internet.

10. Finally, this appendix analyses some recent and near-future developments in tracking. It explores the likely impacts of Google’s recent announcement of its intention to end support for third-party cookies in Chrome, and other proposals within the web standards community to use client-side privacy-enhancing technologies (PETs) to limit tracking by shifting a significant proportion of the data processing to the device itself. Some proposals for privacy-enhancing behavioural targeting and retargeting have been put forward for discussion within standard setting forums and the broader digital advertising community, and could potentially allow advertisers to preserve some of the current ability to target audiences based on their interests and intent. Proposals for privacy-preserving attribution have also been put forward. Most of the proposals so far focus on browsers and are not directly applicable to advertising on mobile apps. There is a risk that these technologies might at least initially reduce revenues for ad-funded publishers, as well as further entrenching the position of large ‘walled-garden’ platforms with alternative means of identifying users across the web.
11. Considering the issues raised in this appendix, we conclude with some areas for potential further work in collaboration with the ICO on consumer protection.

## **Introduction**

12. This appendix describes the tracking of internet users and their devices for personalised advertising.
  - (a) Tracking is the collection and linking of data about a user’s or device’s activity across multiple websites or applications, and the retention, use or sharing of that data. The goal of tracking is to link together the activity of a single user across different browsing sessions, properties (webpages and apps) and devices, to build a more complete profile of the user that could, among other things, help deliver personalised advertising, evaluate the effectiveness of advertising on conversions, and inform advertisers’ spending decisions.

- (b) Much of this data is not personal data in and of itself, but can become personal data when it can be aggregated and combined so that it can be related to an identifiable person.
  - (c) Unlike contextual advertising, which relies on information about the content and context of the webpage or app that the user is currently viewing (such as keywords or topics), personalised advertising is concerned with knowing about an individual to determine whether to show them any ads, which ads to show them, and to measure their behaviour after they were exposed to the ad. This may include their demographic data or interests, whether volunteered or inferred, as well as past browsing and purchasing behaviour. In order to build a comprehensive individualised profile, users may be tracked across multiple channels (websites, apps, different devices, times and locations). The role of data in digital advertising is the subject of Appendix F.
- 13. Personalised advertising is currently dependent on tracking, with the best-known method of tracking being third-party cookies. However, there are many other methods of tracking, such as fingerprinting and embedding third-party code in first-party applications via pixels, tags and Software Development Kits (SDKs), or directly matching and sharing identifiers between companies. This appendix discusses the mechanisms of tracking, covering web, mobile and cross-device tracking as well as how identification is achieved more generally. Following this, we describe tracking in adtech, and explore the relationship between tracking and market power. Finally, we assess users' control (or lack thereof) over whether they are tracked and highlight new technologies that can better protect users and their data by default.

## **Tracking – identification and technologies**

- 14. This section describes how tracking works from a technical point of view. This includes an overview of the core concepts of identification, and how tracking is implemented in practice in a variety of both web and mobile technologies.
- 15. There are two main reasons why understanding the mechanisms of tracking is important:
  - (a) First, in order to assess to what extent users have visibility and control over the information that businesses collect and infer about them. This is further discussed in Appendix L and Chapter 4; and
  - (b) Second, in order to assess whether data collection may be easier for certain platforms that have access to more accurate methods of tracking and more opportunities to track. If this is the case, platforms that can

better track users have an advantage over competitors in their ability to provide personalised advertising services.

## **Identification**

16. Identification is about distinguishing individuals (or telling them apart) from other individuals and recognising the same individual over time. This section discusses identification as a cornerstone of tracking, and explains how identification is achieved, the omnipresence of identifiers (or identifying information) in our daily lives both online and offline, and how identifiers can be better or worse at identifying. We discuss tracking as the practice of linking data from various contexts to a single person, often represented by an identity graph, and how identifiers themselves may not look like personal data, but the linking and combining of identifiers enables the aggregation of large datasets of personal data about individuals.
17. Although advertisers are primarily interested in identifying individual people or users, identification of objects that are closely associated with users such as their laptop, mobile device or their instance of a web browser is often close enough to individual identification. This is because devices are quite personal: most people do not use multiple browsers per device and use at most a handful of devices. Indeed, the ICO identifies a 'device fingerprint' and information which relates to a device a user is using as personal data.<sup>2</sup> Unique device identifiers (UDIDs) are personal: people often don't share their phone or laptop, so knowing the UDID for a person's device can potentially allow a lot of their observed activity to be attributed to that person.

## **Identifiers**

18. Identifiers are pieces of data which help identify an individual, or their device. Many kinds of data can potentially be used as identifiers. Potential identifiers are described as 'strong' or 'weak' depending on how useful they are at helping to distinguish or single out an individual from other individuals. Strong identifiers are a) **unique**, allowing that individual to be precisely singled out from others; b) **persistent**, allowing that individual to be recognised across time; and c) **available**, so that they can be accessed and used. The strength of potential identifiers also depends on context. This context includes, but is not limited to, the extent to which the identifier can be linked to other identifiers.
19. Identifiers that are weak on their own may be combined into a strong identifier. For instance, a person's name on its own may be a weak identifier,

---

<sup>2</sup> [ICO Guidance: What are identifiers.](#)

especially if it is a common name shared with many others. Similarly, a person’s home postcode on its own is shared with others living in that postcode area. But the combination of their name and their postcode may be enough to uniquely identify that person. The strength of any potential identifier depends on attributes of the identifier itself (like uniqueness) in addition to the context (what else it can be linked to it).

20. Identifiers do not need to store information or contain meaning, nor do they need to be interpretable by humans, in order to identify an individual. A string of digits, like a mobile advertising ID (MAID), may look meaningless to a layperson without context and the ability to link it to other information about that individual. But to many adtech providers in the context of the current adtech ecosystem, MAIDs are very strong identifiers. Similarly, the context and capabilities of data controllers and processors are important in assessing whether data can be considered anonymised or merely pseudonymised. For data to be anonymised, the data cannot reasonably<sup>3</sup> be linked back to or re-identify individuals or their devices.<sup>4</sup>
21. Table G.1 is a non-comprehensive list of examples of web and mobile identifiers, and how unique, persistent and available they are. We discuss each identifier in turn below.

**Table G.1: Some identifiers and their uniqueness, persistence and availability**

<i>Identifier</i>	<i>Unique</i>	<i>Persistent</i>	<i>Available</i>
Email address	Yes	Yes, unless/until the user changes it.	Only available when given freely by a user, but readily available from data brokers as often given freely.
Phone number	Yes	Until user changes it. Users often keep their number even when changing devices. <sup>5</sup>	Only visible to apps with special permissions, <sup>6</sup> but readily available from data brokers. <sup>7</sup>
Internal IDs (eg Account ID)	Yes	Yes.	Only within the company that sets it, unless the company chooses to share it.
Cookies	Yes	Until user deletes them.	In some browsers.
localStorage ('super cookies')	Yes	Until user clears it.	Only available in iFrames. Can be blocked by tracker blockers.
Advertising ID (IDFA on iOS, AAID on Android)	Yes	Until user resets it.	Yes, to all apps.
IP address	Yes	On the network. May persist for days, weeks or months. The average persistence in the UK is 18 days. <sup>8</sup>	Always.
IMEI/IMSI	Yes	Yes.	Only visible to apps with special permissions.
MAC address	Yes	Sometimes. Newer iOS and Android devices randomise, but older and other types of devices (eg laptops) do not.	Only visible to apps with special permissions. Assuming the device is not in airplane mode.

<sup>3</sup> Using any reasonably available means, including external data sources.

<sup>4</sup> See [ICO guidance on Anonymisation](#), which at the time of writing is undergoing an update.

<sup>5</sup> Mobile number portability has been adopted in many countries since 1999. The UK established the PAC code for this, regulated by Ofcom, in 2003. Source: [Wikipedia](#).

<sup>6</sup> Requires user consent on [Android](#) and iOS is [not possible at all](#) since iOS 4 (a user must give it themselves)

<sup>7</sup> Often available just by using people searches such as those that [Lifewire](#) lists.

<sup>8</sup> Mozilla and Inria (2020) *Don't count me out: On the relevance of IP addresses in the tracking ecosystem*. Available [here](#).

### *Email address and phone numbers*

22. Email addresses and phone numbers are some of the oldest and most familiar identifiers. They are typically given out quite freely by users signing up to online services, so they are often available. When combined with other identifiers they can be useful for identification purposes: they are unique, often associated with just one person (although some people have multiple email addresses and phone numbers, and may have shared email accounts with others), and quite persistent (as people tend to keep their 'main' email address and phone number for a long time). Furthermore, some people provide the same email address to multiple online services upon registration, and to many companies when asked for contact information, signing up for loyalty schemes or subscribing to mailing. Companies often use email addresses to match records (see sections on 'Remarketing lists and Customer Match' and 'Facebook Customer Audience and Offline Conversion' below for examples of this in practice).

### *User account IDs and other internal IDs*

23. Internal IDs are set by companies to group a user's activity. The clearest example is a user account ID, which is a unique ID created by the company on the user's behalf when they first sign up or register, and is joined with the sign-up information (usually an email and password at minimum). Subsequently, the company can associate all data it has on the user with this internal account ID (unless there are internal data sharing/linking restrictions). Companies may set internal IDs for different products or categories of user behaviour. Typically, companies do not share internal IDs outside their systems, but this may not always be the case. As users are often required to log into their accounts (authentication) to access certain services, user account IDs are often an effective way to achieve cross-device identification (discussed further in section below on 'Linking identifiers, identity resolution and cross-device tracking').

### *Cookies*

24. Cookies are small text files that a website puts into the browser when a user visits that website. When a user visits a website, the website checks if the browser already contains a cookie they set previously for the user. If not, the website puts a new cookie in the browser. The cookie may contain data (such as the user's login status) or just a string of letters and numbers to serve as an identifier. In subsequent visits to the website, the browser sends the

cookies of the user to the site. This enables the site to recognise and identify users and build a browsing history on the user over time.

25. Cookies can be set<sup>9</sup> by first and third parties. The web standards community generally defines a cookie as first party when the registrable domain<sup>10</sup> of the page visited by the user matches the registrable domain of the cookie. If the registrable domains do not match, then the cookie is considered third-party to the page. To illustrate, facebook.com can set a first-party cookie on a browser that is visiting a webpage on facebook.com, and facebook.com can set a third-party cookie on a browser visiting guardian.co.uk. The first party is the site the user is visiting, which changes as they browse; thus the same cookie may be first or third party depending on the user's context. There is no technical difference between how first and third-party cookies work intrinsically, although browsers may treat them differently.<sup>11</sup>
26. Traditionally, first parties usually set cookies for reasons related to providing a service to the user (such as remembering their shopping cart items), whereas a third-party may set cookies to track a user across sites, often for personalised advertising and measurement and attribution services. Thus, third-party cookies are often called tracking cookies, although as discussed further below, the distinction between first and third-party cookies is increasingly blurring due to the practice of first parties adopting third-party code and cookies.
27. Third-party cookies are a well-established method of cross-site tracking. Trackers can read cookies that they set across multiple websites. To illustrate, if two websites A and B both allow the same tracker E to set and read third-party cookies, E can set a third-party cookie on the user's browser when they visit A, and E can retrieve that cookie when the user visits B and recognise that user from their previous visit to A. This allows the tracker to link the sites a user has visited from a browser together.
28. To maintain web security, modern browsers will only send a cookie if the cookie's domain matches the domain requesting it - in other words, browsers adopt a cookie same-origin policy. Therefore, cookies are specific to and can only be accessed by the domain that set them, and domains cannot directly read cookies set by another domain. However, adtech intermediaries often have incentives to establish a common identifier for users so that they can

---

<sup>9</sup> For a technical discussion of how cookies are set, see the section on HTTP and JavaScript under Tracking Technologies section later in this appendix.

<sup>10</sup> Registrable domain is effective top-level domain plus one additional label (eTLD+1). For instance, 'www.google.com' and 'news.google.com' share the same registrable domain.

<sup>11</sup> For example, Chrome is changing to a default where cookies can only be read when the domain in the URL is equal to the cookie's domain (SameSite attribute changes, due to come into force in July 2020). SameSite is discussed in the section on browser innovations under user control later in this appendix.



share information about them. To do this, they may engage in cookie matching (discussed later) as a way of circumventing the cookie same-origin policy.<sup>12</sup>

29. Third-party cookies are not perfectly persistent, as users are able to manually block or clear them. Major browsers, such as Safari<sup>13</sup> and Firefox<sup>14</sup>, have recently started to block them by default. Chrome has announced that it plans to phase out support for third-party cookies within two years.<sup>15</sup>
30. However, as we will discuss, there are many alternatives to third-party cookies for cross-site tracking.<sup>16</sup>

### *Local Storage*

31. localStorage is a Web API<sup>17</sup> and part of the Web Storage API<sup>18</sup> which works similarly to cookies, but is less well-known and thus less likely to be deleted/blocked by a user. It is client-side only, meaning the data added to it stays on the browser and does not get transferred to the website's web server (although websites can read data from it). localStorage can store more data than cookies (up to 5MB). It stores data that websites want to save as text in key value form like cookies. However, localStorage stays in the browser for a longer period than cookies do on average (until the user clears it), beyond the session of a user (whereas only some cookies do) making it more persistent in practice than most cookies.<sup>19</sup>

### *Mobile Advertising IDs (MAIDs)*

32. Mobile advertising IDs (MAIDs) are strings of alphanumeric characters assigned to mobile devices. On iOS devices they are called IDFA (Identifier for Advertisers), and on Android devices they are called AAID (Android Advertising ID)<sup>20</sup>. MAIDs are unique, mostly persistent, and available to all mobile apps (without the need for user permission) and advertisers who embed code in those apps.

---

<sup>12</sup> Cookie matching is often needed for different adtech participants to share information about users' browsing activity across different websites. By contrast, app developers do not need to match different identifiers to share information about users' activity on mobile apps. This is because iPhones and Android phones have a device-wide unique identifier that is shared to all apps for the purposes of in-app mobile advertising, which discussed in the section below on 'Mobile Advertising IDs (MAIDs)'.

<sup>13</sup> WebKit, [Intelligent Tracking Prevention](#). (WebKit is the browser engine Safari is based on.)

<sup>14</sup> Mozilla, [Enhanced Tracking Protection](#).

<sup>15</sup> Chromium Blog, 'Building a more private web: A path towards making third party cookies obsolete', available [here](#).

<sup>16</sup> Including but not limited to DOM storage, cached scripts, localStorage, HTTP request headers and parameters.

<sup>17</sup> APIs are software developers use that let one program access and interact with another program. Common Web APIs are listed [here](#).

<sup>18</sup> More information on the Web Storage API is available [here](#).

<sup>19</sup> This [page](#) has an accessible explanation for Local Storage and its shortcomings.

<sup>20</sup> Sometimes called the AdID. In this appendix we use AAID to describe Android's advertising ID.

33. MAIDs were created by iOS and Android in 2013 and are not required for any essential device or networking functionality. They are used by advertisers and app publishers seeking to monetise their apps via personalised advertising. MAIDs are a bedrock for personalised advertising on mobile, and are strong identifiers in widespread use in the adtech ecosystem, playing a similar role to cookies (although there is no need to match them across apps because they are device-wide identifiers).
34. MAIDs may be less well-known than cookies among consumers and end-users. If a user were aware and wanted to turn off tracking via their MAID, they can only do this on iOS, by turning on 'Limit Ad Tracking' which sets the IDFA to a string of zeros (which if enough others also do, renders it non-unique).<sup>21</sup> On Android, users can reset the AAID, but are immediately given another one, so must in practice regularly reset it to avoid it being used by trackers. There is no way to completely turn off an AAID. There does not appear to be any technical reason why Google could not implement the same feature as Apple's IDFA to enable the user to set the AAID to a string of zeros. Google told us that it does not link subsequent AAIDs to old ones, but we note that this does not preclude apps,<sup>22</sup> OEMs or MNOs from doing so.

#### *IP address*

35. IP addresses are dynamically assigned by the network interface when a user connects to a network (eg Wi-Fi or 4G). They are sent in every request a user makes over the internet, which happens at least once (usually multiple times) when they visit a website. IPs are essential to how the internet was designed to work, for communication and data transfer between devices. They identify the client on the network and are used to prevent fraud/spam. Availability and uniqueness are essential for ensuring IP packets travelling as part of the same request can be reconciled.<sup>23</sup> IPs aren't very persistent, as they change when you go to a new network. Used in conjunction with other identifiers, IPs can be useful for tracking, especially cross-device tracking. IPs can be hidden by sophisticated users who use trusted VPNs<sup>24</sup> or Tor.

---

<sup>21</sup> See Apple's documentation for iOS apps on the IDFA ([here](#)): 'when ad tracking is limited, the value of the advertising identifier is 00000000-0000-0000-0000-000000000000'.

<sup>22</sup> However, in their Developer Guidelines (available [here](#)) Google does ask app developers not to link the new AAID of users who have recently reset AAID to these users' old AAID.

<sup>23</sup> HTTP requests are discussed in the section on browser functions.

<sup>24</sup> We note that there are many allegations that some VPN providers (especially those that are free at point of use) sell data about their users. In 2013, Facebook acquired a VPN provider, Onavo, and allegedly used it to collect data about users and to monitor usage of competitors' apps. (See, eg, 'Apple says it's banning Facebook's research app that collects users' personal information', available [here](#).)

### *IMEI/IMSI*

36. These are strong identifiers that uniquely identify a mobile device (IMEI) and SIM (IMSI).<sup>25</sup> They are not disclosed to apps by mobile operating systems unless they have special permission, as we discuss later in the section on permissions models in mobile. The IMSI is shared with the user's mobile network operator (MNO) whenever they connect to a nearby cell tower and can be used to track location. In the US, MNOs have been found selling this data (location paired with IMSI).<sup>26</sup>

### *MAC address*

37. These are hardware identifiers in every internet-enabled device. The device sends pings out constantly, to find nearby Wi-Fi or Bluetooth devices. These can be intercepted by wireless beacons nearby. Beacons are set up at events or businesses and are good for short distance location tracking and building social proximity graphs.
38. MAC addresses are now routinely randomised on both Android and iOS, but not laptops, tablets or other internet-enabled devices.

### *Linking identifiers, identity resolution and cross-device tracking*

39. There are two main reasons to link identifiers and create 'identity graphs'<sup>27</sup>:
- (a) Identity graphs combine many identifiers and thus make a collection of weak identifiers more persistent and unique.
  - (b) To build as comprehensive picture as possible of an individual, including across different contexts and devices (ie a cross-device graph).
40. Weak identifiers on their own are not reliable for identification.
- (a) Many weak identifiers are not persistent. For example, IP addresses may be changed by users disconnecting their modem, and cookies can be deleted by users. Linking weak identifiers helps trackers to re-establish identification if one or more (but not all) of the identifiers in the profile change. For example, if a user deletes cookies but retains the same IP address, trackers could observe that the new cookies are being read from

---

<sup>25</sup> IMEI stands for International Mobile Equipment Identity. IMSI stands for International Mobile Subscriber Identity.

<sup>26</sup> The FCC [have fined](#) some MNOs for this conduct, which the FTC also investigated, following [Motherboard reporting](#) it.

<sup>27</sup> An identity graph for a user is a list of identifiers for that user structured in a [graph](#), where edges (connections) represent a deterministic (used together) or probabilistic connection (share another attribute/identifier, such as timestamp) between the identifiers.

a browser using a known IP address, and infer that the new cookies should be linked to the old cookies on the user's identity graph.<sup>28</sup>

(b) Weak identifiers on their own may not be sufficiently unique, but might allow identification of individuals or their device/browser in combination with other identifiers. This practice is known as fingerprinting, and is discussed below.

41. An individual can have many identifiers, and each of these may be associated with different information about them. Correctly linking multiple identifiers allows advertisers to bring different information together to form a more complete picture of them. Identifiers can be linked across many dimensions, including different devices, different datasets, different websites and apps, across locations, and across time.
42. This process of linking together multiple identifiers across different dimensions to build a single unified profile for individuals is often known in adtech as 'identity resolution'. There are adtech providers that specialise in identity resolution services, attempting to match and connect identifiers into unified customer profiles at scale. They license or provide access to identity graphs to other market participants. (These are discussed in more detail in the section below on 'Data management platforms and data brokers'.)
43. The most extensive and useful identity graphs are typically maintained by platforms with access to rich sources of user data (such as Google and Facebook), DMPs, CDPs,<sup>29</sup> and other specialised adtech services.
44. Cross-device linking is particularly valued by advertisers. Advertisers can build richer profiles of individuals and target users more effectively if they can aggregate data across a user's devices. Also, cross-device graphs are very valuable for frequency capping and to measure advertising effectiveness, as it may be common for individuals to view ads, research, and make purchases on different devices.
45. Advertisers, publishers and adtech providers can make trade-offs between certainty and scale in identity resolution. It is helpful to think of two broad kinds of identifier matching along a continuum:

---

<sup>28</sup> This basic idea is applied in a tracking technology called the 'Evercookie'. The documentation for it states: "Evercookie is a JavaScript API that produces extremely persistent cookies in a browser. Its goal is to identify a client even after they've removed standard cookies, Flash cookies (Local Shared Objects or LSOs), and others. This is accomplished by storing the cookie data on as many browser storage mechanisms as possible. If cookie data is removed from any of the storage mechanisms, Evercookie aggressively re-creates it in each mechanism as long as one is still intact."

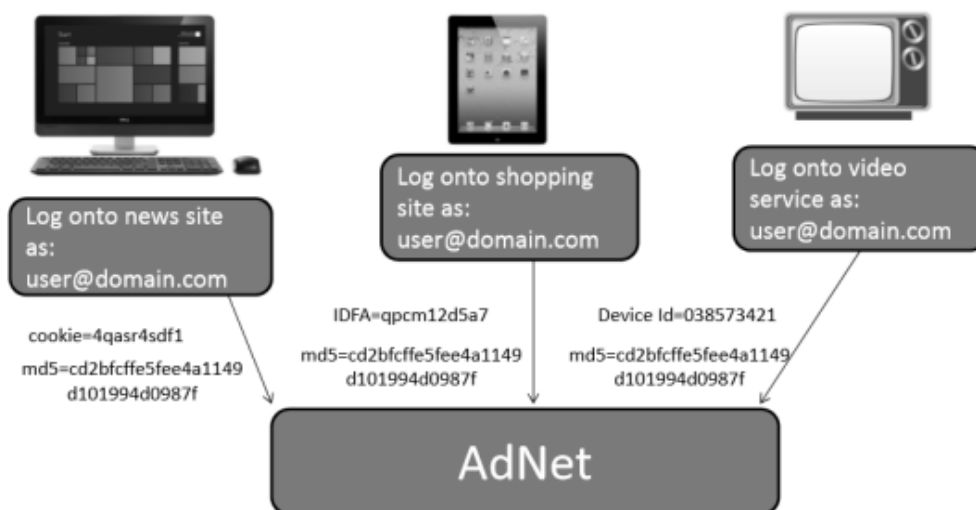
<sup>29</sup> Customer Data Platform, as described [here](#).

- (a) **Deterministic identifier matching** is when two identifiers can be linked with confidence they belong to the same individual. For example, if an individual provides the same email to an advertiser and when signing up to Facebook, the advertiser can send the (hashed<sup>30</sup>) email to Facebook and Facebook can look for an exact match. (This is discussed in the section below on Facebook Custom Audience.)

Similarly, if a user logs into their Facebook account using their desktop browser and also using an app on their mobile device, Facebook can link with certainty its cookie on the user's browser with the user's MAID (ie cross-device tracking) as these identifiers co-occurred with the internal login ID/details. Deterministic matching is good for accuracy, but not scale, as much online activity does not require a user to provide strong identifiers.

Figure G.1 illustrates a scenario where multiple identifiers for different devices can be linked together by the user's email.

**Figure G.1: Deterministic linking of different identifiers via another identifier (email).**



Source: FTC (2017) *Cross-device tracking: measurement and disclosures*. Available [here](#).

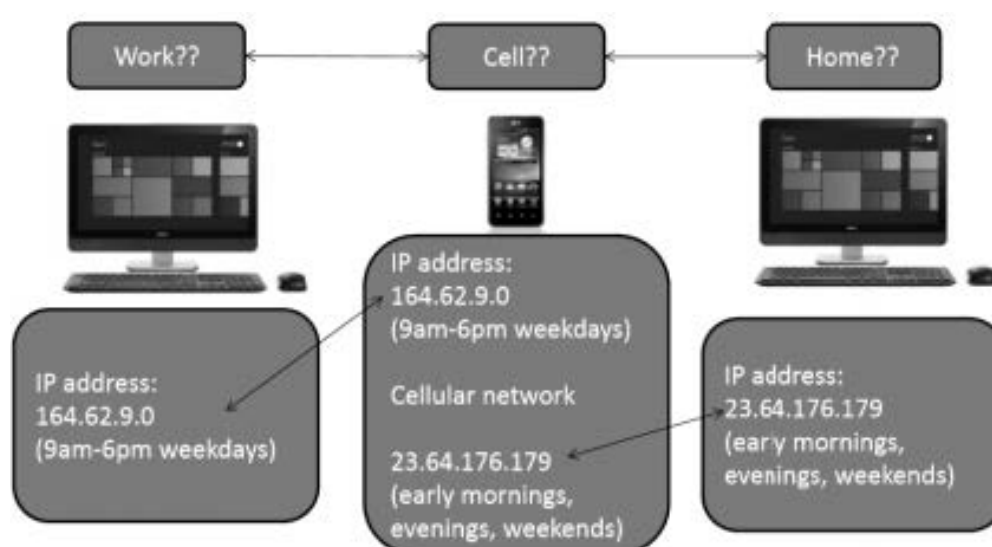
- (b) **Probabilistic identifier matching** uses multiple weak identifiers (such as IP addresses, or browser types) and other regularly occurring signals as inputs to a statistical model. For example, a tracker may observe that

<sup>30</sup> A hash function is a one-way mathematical function that converts any amount of data into a string of letters and numbers (a hash value) with a fixed length. Hash functions are designed so that hash values cannot be inverted to find the original input data. However, hashing algorithms are designed to avoid 'collision' (two different inputs leading to one output), so if two hashes match then it is highly likely that their inputs were the same too, allowing them to enable linking datasets together. It is important to note that hashing does not solve all privacy concerns – a hashed identifier can nevertheless be a persistent, unique identifier that allows linking a person across databases, devices, and contexts.

three devices have overlapping IP addresses at certain times and may infer that there is a high probability that the three devices belong to the same person. Figure G.2 illustrates.

This process does not always deliver fully accurate matches; for example, many people may use the same IP address or network as in a coffee shop. In these cases, adding further signals like browsing history or using device proximity techniques<sup>31</sup> can improve estimates. Estimates of the accuracy of cross-device probabilistic linking can be as high as 97%.<sup>32</sup> Sometimes hundreds of identifiers are used in probabilistic matching.<sup>33</sup> Probabilistic matching sacrifices some accuracy to achieve greater scale.

**Figure G.2: Probabilistic matching cross-device by correlating time and IP address**



Source: FTC (2017) *Cross-device tracking: measurement and disclosures*. Available [here](#).

## Fingerprinting

46. Fingerprinting has many similarities to probabilistic matching. The key idea of fingerprinting is that pieces of information which are weak identifiers on their own (such as screen size and colour depth, system fonts, and time zone) can be combined into a 'fingerprint', a strong identifier that uniquely identifies an individual browser or device. Many kinds of information can be used to make a fingerprint, including a user's unique pattern of mouse movement or scrolling or the way that a user holds their device (using data from many sensors on our devices).<sup>34</sup>

<sup>31</sup> For instance, near-field communication (NFC) sends data from one device nearby to another. Similarly, beacons and Bluetooth are commonly used in proximity marketing.

<sup>32</sup> Digiday, 'Cross-device tracking, explained', available [here](#).

<sup>33</sup> See Admonsters, 'Probabilistic Identifiers and the Problem with ID Matching', available [here](#).

<sup>34</sup> See for example Behavioural Biometrics by BioCatch, available [here](#).




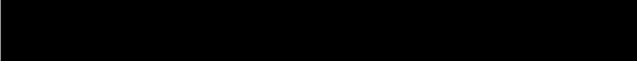
47. To illustrate fingerprinting, Figure G.3 shows an example output of the educational tool Panopticlick.<sup>35</sup> Each item in the list of browser characteristics contains different amounts of information.<sup>36</sup> No item on its own is enough to uniquely identify a browser install, but the combination of them can, as it is very unlikely for other browser installs to share the exact same configuration.

---

<sup>35</sup> [Panopticlick](#) – a free tool from the Electronic Frontier Foundation.

<sup>36</sup> Bits are the units of entropy and self-information, which are measures of information content. To illustrate very briefly, suppose an identifier X can only take one of two values (A or B) with equal probability (0.5). If we learn for an individual that the value of the identifier is A, then the 'self-information' of this particular outcome is 1 bit ( $-\log_2 0.5$ ). The entropy is the expected value of the self-information of all possible outcomes and indicates how 'informative' or 'surprising' learning the value of that identifier would be on average. In this case, the entropy of this identifier X is 1 bit. 33 bits of identifying information would be enough to uniquely identify a single person out of 7.8 billion people ( $2^{33} = 8.5$  billion). In practice, the amount of entropy of an identifier depends on context and what else is already known. If an individual's postcode is known, the added information of their city gives no additional information. Entropy is useful for understanding various privacy budget proposals being developed, including by Google in its Privacy Sandbox, which are discussed later.

**Figure G.3: A browser fingerprint**

Browser Characteristic	bits of identifying information	one in <i>x</i> browsers have this value	value
User Agent	5.6	48.34	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.138 Safari/537.36
HTTP_ACCEPT Headers	2.74	6.69	text/html, */*; q=0.01 gzip, deflate, br en-US,en;q=0.9
Browser Plugin Details	2.99	7.95	Plugin 0: Chrome PDF Plugin; Portable Document Format; internal-pdf-viewer; (Portable Document Format; application/x-google-chrome-pdf; pdf). Plugin 1: Chrome PDF Viewer; ; mhjfbmdgcfjbbpaeojofohoefgiehjai; (; application/pdf; pdf). Plugin 2: Native Client; ; internal-nacl-plugin; (Native Client Executable; application/x-nacl; ) (Portable Native Client Executable; application/x-pnacl; ).
Time Zone Offset	3.87	14.63	-60
Time Zone	4.05	16.62	Europe/London
Screen Size and Color Depth	5.26	38.39	2560x1440x24
System Fonts	4.04	16.41	
Are Cookies Enabled?	0.2	1.15	Yes
Limited supercookie test	1.53	2.89	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No, openDatabase: true, indexed db: true
Hash of canvas fingerprint	6.78	110.24	
Hash of WebGL fingerprint	7.29	156.57	
WebGL Vendor & Renderer	6.3	78.55	
DNT Header Enabled?	1.03	2.05	False
Language	0.92	1.89	en-US
Platform	1.17	2.25	Win32
Touch Support	5.04	32.83	Max touchpoints: 10; TouchEvent supported: false; onTouchStart supported: false
Ad Blocker Used	0.37	1.3	False
AudioContext fingerprint	2.44	5.42	124.04344884395687
CPU Class	0.16	1.12	N/A
Hardware Concurrency	1.86	3.64	4
Device Memory (GB)	2.46	5.5	8

Source: EFF Panopticklick tool, Available [here](#).



48. At a technical level fingerprinting can be done regardless of users' privacy preferences.<sup>37</sup> Users cannot decline to be fingerprinted due to the wide surface area of data sources used, and the practice is very difficult completely prevent. Browsers are constantly innovating in an arms-race against trackers to make fingerprinting more difficult and less effective.<sup>38</sup>
49. Many of the browser and system characteristics used for fingerprinting cannot be modified by the user. For instance:
- (a) Many of the characteristics are determined by the user's hardware (such as screen size and colour depth, touch support, device memory, etc.).
  - (b) The HTTP header User-Agent<sup>39</sup> is routinely sent and collected as part of HTTP requests (see section below on 'Hypertext Transfer Protocol'), but it can also reveal information which can be used for fingerprinting, including by adtech providers. User-Agent has many legitimate uses, and there is an active discussion within the web standards community about how these functions may be fulfilled in a more privacy-preserving way,<sup>40</sup> and about the impact that deprecating User-Agent would have on websites' functionality and their access to data about their users.<sup>41</sup>
  - (c) Similarly, Canvas and WebGL fingerprinting exploit the fact that websites can give a user's browser a task to render an image using a predefined script. The image can contain multiple elements, such as lines, colours, gradients, shapes, text, etc. Different devices will draw the image in a slightly different way,<sup>42</sup> with small variations and differences that are not noticeable to the human eye, but which can be detected by the tracker. The image output is hashed and sent back to the tracker.

---

<sup>37</sup> 'Opinion 9/2014 on the application of Directive 2002/58/EC to device fingerprinting', provides that 'device fingerprinting for the purpose of targeted advertising requires the consent of the user' (7.2, page 9), see also the ICO's guidance on fingerprinting in 'What are the rules on cookies and similar technologies'.

<sup>38</sup> Narayanan (2018) *Against privacy defeatism. Why browsers can still stop fingerprinting*. Available [here](#).

<sup>39</sup> The User-Agent HTTP header is a string that tells the server the type and version of the client's browser and device. It is sent with every HTTP request made by the browser (at least once when a user visits a webpage).

<sup>40</sup> As part of their effort to combat fingerprinting, Safari decided to freeze the User-Agent string (see [here](#)), so that it will no longer change in the future and create variation that be used for fingerprinting. At the time of writing, there is currently a proposal by Chromium to deprecate the User-Agent string and replace it with [User Agent Client Hints \(UA-CH\)](#). See this explainer (available [here](#)) for a list of the current uses for the User-Agent string, and how User Agent Client Hints might preserve these functionalities whilst revealing less information that could be used to fingerprint for persistent user tracking.

<sup>41</sup> See, for instance, this [debate](#) on GitHub about Chromium's proposal to follow Safari and (partially) freeze the User-Agent string.

<sup>42</sup> These differences could be due to differences in font rendering, smoothing, anti-aliasing, as well as other device features.

50. Many of the attributes used in fingerprinting are information that browsers cannot help sending to websites and are core to how the internet currently works.<sup>43</sup>
51. Unlike cookies, fingerprinting does not require any data to be stored on the user's computer (ie fingerprinting is 'stateless'), so users cannot delete or reset anything to break identification short of switching browsers and devices. Fingerprinting works even if users browse in incognito or private modes.
52. Ironically, sometimes taking active measures against tracking can make a person more vulnerable to fingerprinting, because it makes their configuration more unusual. For example, installing browser extensions or plugins to block trackers could make a user's browser more identifiable, if relatively few other people have these extensions installed.<sup>44</sup> In theory, switching to a browser with a low market share without taking active measures to obfuscate its identity, such as modifying the User-Agent string to pretend to be another browser (browser spoofing),<sup>45</sup> could also increase a user's vulnerability to fingerprinting.<sup>46</sup>
53. All of these factors make fingerprinting robust and hard to prevent. They also make it difficult to detect, since the identifying information used in fingerprinting is part of how the internet currently works and is requested by websites that don't engage in fingerprinting (eg IP addresses). Some browsers, like Tor, are designed to counter fingerprinting by making their users look like each other. This mitigation strategy is known as uniformity. The other principal strategy against fingerprinting is randomization. Tor,<sup>47</sup> Safari<sup>48</sup>, Firefox,<sup>49</sup> and Brave<sup>50</sup> have all published implementation recommendations on combating fingerprinting. The W3C published guidance on mitigating browser fingerprinting,<sup>51</sup> and several browser plugins exist that aim to counter fingerprinting.

---

<sup>43</sup> For instance, fingerprinting based on metadata from a user's network. See Network-Based Website Fingerprinting from the IETF available [here](#).

<sup>44</sup> This is one of the reasons why Tor considers plugins to be the most severe fingerprinting threat. (See 'The Design and Implementation of the Tor Browser', section on Specific Fingerprinting Defenses in the Tor Browser, available [here](#).) In addition to their very presence adding entropy to fingerprints, plugins are also capable of extracting information beyond what browsers normally provide to websites, and can be used to store unique identifiers that are more difficult to clear than standard cookies.

<sup>45</sup> See [History of the browser user-agent string](#) to see how browsers pretended to be one another, in order to manipulate websites and apps that use browser sniffing (or browser detection) to serve different content depending on the user's browser.

<sup>46</sup> An experiment by the Electronic Frontier Foundation (available [here](#)) found that the User-Agent string for the average browser contains 10.5 bits of identifying information, which means that on average only one person in about 1,500 will have the same User-Agent string as another person.

<sup>47</sup> Tor's [design specification](#) for their browser covers this in depth from a basis of principles such as unlinkability.

<sup>48</sup> WebKit's guidance from 2011 on how clients can watch out for vectors of fingerprinting. Available [here](#)

<sup>49</sup> Mozilla have an anti-fingerprinting project ongoing in conjunction with Tor Uplift. Details [here](#).

<sup>50</sup> See which fingerprinting vectors Brave block [here](#) and limitations of block lists.

<sup>51</sup> See the W3C fingerprinting guidance available [here](#).

54. Fingerprinting is straightforward from a technical perspective. In principle any website that can run JavaScript can perform browser fingerprinting. Estimates of the prevalence of unambiguous fingerprinting techniques (like canvas fingerprinting) being used suggest that it is relatively uncommon.
- (a) Englehardt and Narayanan (2016)<sup>52</sup> measured the top 1 million websites and found 14,371 (0.14%) websites performed canvas fingerprinting, and that more popular sites are more likely to have fingerprinting scripts. (For instance, of the top 1,000 websites, 5.10% had canvas fingerprinting scripts.)
- (b) Das et al. (2018) measured smartphone sensor API usage on the top 100K websites, found that 3,695 (3.70%) websites had scripts to access sensors (such as motion and orientation sensors). The authors also found 1,991 websites (1.99%) had canvas fingerprinting scripts that also access sensors.<sup>53</sup>
55. These estimates are a lower bound, in the sense that these authors directly analysed the JavaScript of these websites and found code implementing techniques that could not plausibly be for any other purpose. However, many more websites could be covertly using the information that many websites routinely receive, arguably for legitimate purposes (such as User-Agent, screen size and colour depth), for fingerprinting instead. Users, or indeed external researchers, would not be aware of this. Both Google<sup>54</sup> and Mozilla<sup>55</sup> assert that fingerprinting has been increasing over time.
56. There is an active discussion within the web standards community about curtailing browsers' vulnerability to fingerprinting by limiting the amount of information that browsers expose to websites, whilst balancing the need for websites to get access to information in order to provide useful functions, within the framework of a 'privacy budget'. This is discussed in the section below on 'Recent and near-future developments in tracking'.

---

<sup>52</sup> Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1388-1401). Available [here](#).

<sup>53</sup> Das, A., Acar, G., Borisov, N., & Pradeep, A. (2018, January). The Web's Sixth Sense: A Study of Scripts Accessing Smartphone Sensors. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1515-1532). Available [here](#).

<sup>54</sup> Google states: 'Browsers have been making changes to how cookies are treated. Blunt approaches to cookie blocking have been tried, and in response we have seen some user-tracking efforts move underground, employing harder-to-detect methods that subvert cookie controls. These methods, known as 'fingerprinting', rely on various techniques to examine what makes a given user's browser unique.' [Combating Fingerprinting with a Privacy Budget](#).

<sup>55</sup> Mozilla states: 'Despite a near complete agreement between standards bodies and browser vendors that fingerprinting is harmful, its use on the web has steadily increased over the past decade.' Mozilla, 'Firefox 72 blocks third-party fingerprinting resources', available [here](#).

## **Tracking technologies**

57. To understand the challenges in giving users control over their data and to understand how market participants could have competitive advantages from more opportunities to collect data, this section considers how tracking works in practice. In practice, the linking of identifiers occurs in the context of the ecosystems in which they are implemented. This includes both web and mobile devices<sup>56</sup> and the communication networks they use to exchange data.
58. In relation to web communication, a third-party tracker must be able to send data from a user's device back to itself at some point, or at least be able to connect to the end-user's device to view it. The data itself may be user-generated or device-generated. In either event, it is helpful to consider how users interact with first parties (the website or app that the user is using), how trackers 'get in' to this otherwise private interaction and how they 'get out' the data/identifiers they are after. Therefore, communication networks are an important cross-cutting topic as we examine the devices and platforms these networks connect.
59. Desktop and mobile differ in their typical usage. On mobile, users spend a significant amount of their time accessing online services via apps rather than via web browsers, although the relative proportions depends on the app.<sup>57,58</sup> By contrast, although desktop applications are used, by default most of user engagement with online services and ads is on websites accessed via a web browser. Therefore, we focus on two main ecosystems: web browsers in a desktop context, and apps in the mobile operating system context.<sup>59</sup>

### **Web browsers**

60. Users access the internet and online services using web browsers such as Chrome, Safari, Firefox, Edge, Internet Explorer and Brave. Web browsers

---

<sup>56</sup> Other devices, such as the various Internet of Things (IoT), can also be used for tracking, but this appendix does not cover them.

<sup>57</sup> A 2015 report on mobile usage by comScore (available [here](#)) showed that apps accounted for 54% of digital media time spent, but that most time spent on mobile by users is in a few heavily used apps. Further research by Morgan Stanley using comScore data found that, for the top 50 mobile web properties, most properties receive more traffic from mobile browsers than from apps, and that mobile app traffic exceeds mobile browser only for a few heavily used properties. (See Marketing Land, 'Morgan Stanley: No, Apps Aren't Winning. The Mobile Browser Is.', available [here](#).)

<sup>58</sup> A study showing mobile apps are far more privacy invasive to users than mobile browsers. Available [here](#).

<sup>59</sup> We acknowledge that this leaves out digital advertising delivered via mobile web browsers to mobile devices, and digital advertising delivered via desktop applications on desktop devices. On the latter, we note that Windows 10 for instance has a unique advertising identifier which, much like MAIDs, are set on the operating system, and which is on by default and can be accessed by app developers and ad networks (discussed by Microsoft [here](#)). We also leave out tracking and identification for emerging channels of digital advertising, such as: connected TVs (CTV); over-the-top (OTT) video streaming devices (such as games consoles, smart TVs, streaming boxes, internet-enabled smart blue-ray/DVD players, HDMI sticks like Amazon Fire TV stick and Chromecast); digital out-of-home (DOOH) media such as billboards that dynamically display personalised ads to each individual as they walk past; and digital radio ads (which can also be personalised).

have several functions. This section focuses on the functions that facilitate tracking and discuss a range of web tracking technologies, from HTTP headers and parameters to JavaScript, the language of the web and of web trackers in pixels, tags, widgets and iFrames. The key messages of this section are that tracking technologies on the web use features that are core to how the web works, and that these mechanisms are far more diverse and often more robust and harder to avoid than third-party cookie tracking.

### *Browser functions that facilitate tracking*

- *Hypertext Transfer Protocol (HTTP)*

61. Web browsers are designed as portals to the web for users. One of their most important functions is to make requests and receive responses using the Hypertext Transfer Protocol (HTTP), which facilitates communication (the transfer of data) over the web.
62. HTTP is core internet plumbing. When a user visits a web page by entering its URL in a browser or clicking on a hypertext link, the *client* (browser) sends a *request* to a *web server* for its content. The web server sends a *response* to the client which includes the content of the web page and any accompanying scripts (discussed below in section on 'JavaScript').
63. We note several features of HTTP that are currently used to facilitate user tracking:
  - (a) URLs can contain parameters (such as '?key1=value1&key2=value2'), which don't affect the link's destination, but provide a way to pass extra information to the destination site. This practice is known as *link decoration*. These keys and values can be any string of characters and symbols, and are often used for tracking users across different websites.<sup>60</sup> For example, in cookie matching (discussed in more detail in the section below on 'Cookie matching'), adtech providers can decorate links with their IDs for the user.<sup>61</sup>
  - (b) HTTP requests and responses can have metadata that can be used for tracking. For example, *HTTP headers* are fields at the start of a HTTP message that let the client and the server pass additional information with the HTTP request or response. For example, some noteworthy headers include:

---

<sup>60</sup> For an illustration of the use of link decoration in tracking, see Apple's explanation of 'Cross-Site Tracking Via Link Decoration' in WebKit's [Intelligent Tracking Prevention 2.2 update blog](#).

<sup>61</sup> See, for instance, the section on 'How cookie matching works' in Google's Authorized Buyers RTB documentation, available [here](#).

- (i) **Set-Cookie** and **Cookie** – used to instruct the browser to store a cookie and send it back in future requests to the web server.
- (ii) *User-Agent* – a string that tells the server the type and version of the client’s browser and device. This is used by websites to select the most suitable version of a website to return to the client (known as *content negotiation* or *browser sniffing*). For example, websites require information on whether the user is on a desktop or mobile device, in order to scale the webpage appropriately. However, User-Agent is also commonly used for fingerprinting (as discussed in the section above on ‘Fingerprinting’).<sup>62</sup>
- (iii) *Referer* (sic) – the URL address of the previous web page from which a link to the currently requested page was followed. As discussed above, this URL can be decorated and include parameters used for tracking.<sup>63</sup>
- (iv) *ETag* (or entity tag) – intended to be a resource identifier that facilitates caching, which can speed up internet browsing.<sup>64</sup> However, the ETag can also hold arbitrary identifiers, including for the user.<sup>65</sup> ETags could be used as a way of recreating cookies that users have deleted.<sup>66</sup> Modern browsers allow users to clear both cookies and the browser cache using the same interface.

- *JavaScript*

64. Another major function of web browsers is running a JavaScript engine. The JavaScript engine interprets JavaScript (JS), the programming language of

---

<sup>62</sup> As part of their effort to combat fingerprinting, Safari decided to freeze the User-Agent string (see [here](#)), so that it will no longer change in the future and create variation that be used for fingerprinting. At the time of writing, there is currently a proposal by Chromium to deprecate the User-Agent string and replace it with [User Agent Client Hints \(UA-CH\)](#). See this explainer (available [here](#)) for a list of the current uses for the User-Agent string, and how User Agent Client Hints might preserve these functionalities whilst revealing less information that could be used to fingerprint for persistent user tracking.

<sup>63</sup> Browsers are increasingly implementing stricter referrer policies. Safari’s [Intelligent Tracking Prevention 2.3](#) caps the referrer header to the source webpage’s registrable domain (ie eTLD+1), thus removing all link decoration, initially for websites that ITP classified as having cross-site tracking capabilities and now [for all third-party requests](#) and [cross-site document.referrer API requests](#). Firefox has a default ‘strict-origin-when-cross-origin’ referrer policy, and Chrome is intending to implement the same (see [here](#)). (A ‘strict-origin-when-cross-origin’ referrer policy: 1) sends the full path if going from one secure (HTTPS) page to another secure page with the same domain; 2) sends the eTLD+1 if going from HTTPS on one domain to a different HTTPS domain; and 3) sends nothing in the referer header if going from a secure domain to an insecure domain. See [MDN on Referrer-Policy](#).)

<sup>64</sup> Web servers can assign ETags to specific versions of resources at URLs. Clients can cache the current version and tag of the resource. In future requests, the client and web server can just use the ETag to check whether the resource has been updated and, if the locally cached version is still good, the client can use the cached version and save bandwidth.

<sup>65</sup> For instance, a tracker can include the same file (such as a transparent 1x1 pixel image) in every webpage and ensuring that each new visitor is given a different ETag, the ETag would be functionally equivalent to a cookie.

<sup>66</sup> See, for instance, [Evercookies](#).



the web. JS is used to enable dynamic content on websites, but can perform many tasks. We set out some of the powerful features of JS that can facilitate tracking:

- (a) The dynamic nature of JS gives it the ability to get and set attributes on the DOM (a programmatic representation of the webpage),<sup>67</sup> including, for example, the ability to access user input into a form such as their email and password.
- (b) JS has event listeners, which get triggered when a user takes actions such as scrolling or clicking. When a listener is triggered by an event, JavaScript can do something in response, such as send data in a HTTP request, or modify the DOM. Event listeners are used by tags, pixels and other trackers like Google Analytics.<sup>68</sup>
- (c) JS can issue HTTP requests (using the XMLHttpRequest<sup>69</sup> or Fetch API<sup>70</sup>), as discussed above, allowing data as collected from JS features above to be sent to adtech providers and trackers.
- (d) Cookies are available to JavaScript by default (via the document.cookie API).<sup>71</sup> This means that client-side cookie setting and reading, in addition to server-side cookie setting (using the Set-Cookie header in HTTP responses, discussed above), can be done by tracking scripts loaded by the website. In particular, a website loading multiple third-party trackers' scripts could enable multiple third-parties to see all the cookies set by the other scripts, allowing trackers to share and match their cookie ID. Safari described this phenomenon as follows: 'cross-site trackers have started using first-party sites' own cookie jars for the purpose of persistent tracking'.<sup>72</sup> This form of cross-site scripting also introduces vulnerabilities as, for instance, cookies available in document.cookie that are set by one third-party may be stolen by another third-party or attacker.<sup>73</sup>

---

<sup>67</sup> The [Document Object Model](#) (DOM) connects web pages to programming scripts by representing the structure of a document (eg HTML).

<sup>68</sup> Event listeners are parts of JavaScript programs that live on a web page and wait for an 'event' which includes various interactions with a web page that a user might take such as scrolling or clicking. When this happens, the event listener is triggered and stores this fact in the tag/pixel. This is how Google Analytics [works](#).

<sup>69</sup> [XMLHttpRequest](#) is a Web API that allows requests from the client (browser) to a web server to be made dynamically (without the web page being reloaded) as it is executed by the JavaScript engine in the browser.

<sup>70</sup> The [Fetch API](#) is a Web API similar to XMLHttpRequest with some additional functionality. It allows requests from the client (browser) to be made to a server.

<sup>71</sup> MDN, [Document.cookie](#), available [here](#).

<sup>72</sup> For this reason, Safari's [Intelligent Tracking Prevention 2.1](#) deleted cookies created through the document.cookie API (rather than set using HTTP responses) within 7 days, as opposed to the 30-day limit for other cookies in force at the time. Third-party cookies have since been blocked completely.

<sup>73</sup> It is possible to set cookies with the HttpOnly attribute. HttpOnly cookies are not exposed to JavaScript, so tracking scripts on the website cannot read and leak the contents of those cookies. Although the HttpOnly attribute was introduced in 2002, only c.8.31% of Set-Cookie operations use it, according to Chrome telemetry data for July 2018 (reported in this [proposed alternative to cookies](#)).

- *Same Origin Policy*

65. Trackers collect data across multiple websites using JavaScript, cookies and HTTP metadata. One important internet policy that restricts trackers was introduced in 1995: the Same Origin Policy (SOP), which only allows scripts to be run on a webpage if the script comes from that webpage's origin (domain name).<sup>74</sup> The SOP prevents cross-site scripting, which can be used for attacks and for tracking.<sup>75</sup> However, sites can still choose to adopt third-party scripts as their own (giving them the same origin) with the <script> or <img> HTML elements.<sup>76</sup> This is known as cross-site script inclusion and it is how pixels and tags work, discussed in the section below on 'Third-party code in first-party websites'.

- *Browser extensions*

66. Browser extensions are small software modules that users can add to customise their browser with additional features. Chrome, for instance, allows users to download extensions from the Chrome Web Store.<sup>77</sup>

67. Browser extensions (also known as add-ons or plug-ins) are bundles of JavaScript with cross-site permissions, and so they are capable of tracking users. In addition, browser extensions are also capable of extracting information beyond what browsers normally provide to websites, and can be used to store unique identifiers that are more difficult to clear than standard cookies.<sup>78</sup> A 2017 study found 38% of the top 10k Chrome extensions make requests to third parties, and 6.3% leak browsing or search history data whether accidentally or intentionally.<sup>79</sup>

68. Sometimes browsers implement warnings on certain permissions that extensions implement, in order to make consumers aware of just how far reaching their abilities can be. A generic example is shown in Figure G.4. Many sensitive features, such as camera, location and clipboard (copy paste) data may be accessed. The permission policies and defaults on what

---

<sup>74</sup> The SOP has similar intention to the browsers' treatment of requests to read cookies set by different domains (the cookie same-origin policy) discussed earlier.

<sup>75</sup> [This top answer](#) on this question on Security StackExchange explains why the Same Origin Policy is so important.

<sup>76</sup> HTML (Hypertext Markup Language) is the markup notation in which a webpage document is written. Elements are parts of it which contain some content such as text, and they themselves are contained by other elements, forming a tree (the Document Object Model).

<sup>77</sup> [Chrome Web Store](#).

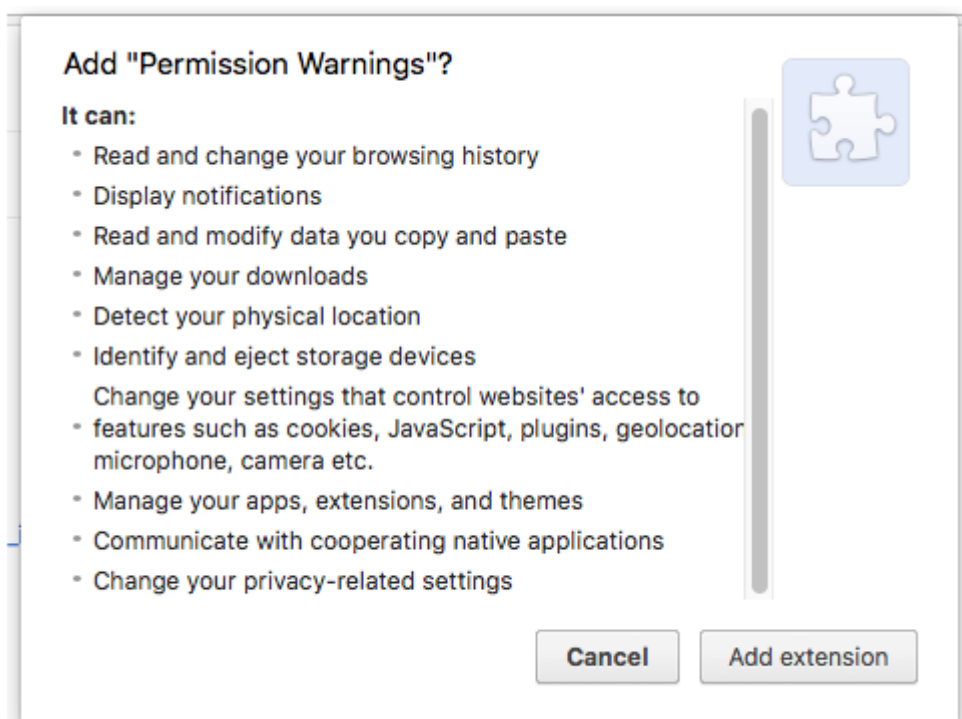
<sup>78</sup> This is one of the reasons why one of Tor's philosophical positions on technology that underpin its design is that 'Plugins must be restricted'. (See 'The Design and Implementation of the Tor Browser', available [here](#).)

<sup>79</sup> Starov, O., & Nikiforakis, N. (2017, April). Extended tracking powers: Measuring the privacy diffusion enabled by browser extensions. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1481-1490). Available [here](#).



extensions can do are varied and may depend on web standards and what leading browsers do in practice.

**Figure G.4: Illustrative example of the permission warnings that explain what a browser extension might be able to do.**



Source: developer documentation from Google available [here](#).

### *Third-party code in first-party websites*

69. In principle, trackers that send user data to third-party domains – that is, domains different from the property that the user intends to visit – would contravene the Same Origin Policy. In practice, most trackers are able to do this by being willingly incorporated into websites by the first party itself. As a result, trackers can be allowed to collect significant amounts of information about users' interactions with the website: at a minimum, the tracker will be aware of the fact that the user visited the site; in some cases, the tracker will be able to observe and extract all information that the website can observe on the user.
70. On websites, third-party trackers take the form of embedded snippets of JavaScript code, often called pixels or tags. This section discusses notable examples, including the Facebook Pixel, Google Analytics, and Google Tag Manager.

71. The Facebook Pixel is embedded on a web page using the `<script>` HTML element.<sup>80</sup> The Pixel is willingly incorporated by website developers who want to measure user interactions with the page to better understand their behaviour, but this data is also sent to Facebook. The Pixel allows Facebook to collect data including HTTP headers, button clicks and other events (eg scrolling), form fields (eg email, passwords a user inputs), the Pixel's unique ID, plus any other data the website wishes to measure (using optional values and custom data events).<sup>81</sup> Facebook requires website developers who incorporate their Pixel to allow them to perform cross-site scripting.<sup>82</sup> The applications of Facebook Pixel for adtech are discussed in the section below on 'Facebook Business Tools'.
72. Google Analytics (GA) has similar functions to the Facebook Pixel, and lets web developers who add GA to their site measure various user interactions such as clicks, sign-ups or whatever else the website wants to measure. This data is sent to Google Analytics. Google Tag Manager (GTM) is a tool used as a container for many tags/pixels, allowing a website to integrate multiple trackers like GA or the Facebook Pixel. The applications of GA and GTM for adtech are discussed in the section below on 'Google Analytics, Floodlight and Google Tag Manager'.
73. Pixels and tags are designed for tracking and analytics, often for marketing and advertising purposes. However, there are other mechanisms which were not primarily designed for tracking but that may nevertheless facilitate tracking. These include, but are not limited to, iFrames and widgets (such as social media like buttons).
- (a) *iFrames* (inline frames) are HTML documents embedded inside another HTML document, akin to a little website within a website. They are often used to embed content into a website from another source. For example, a YouTube video embedded in a news article uses iFrames. iFrames can send data back to the embedded domain's owner – eg an embedded YouTube player can send data back to Google. iFrames may also store identifiers in the browser using third-party cookies, which they can then use to identify users uniquely when they visit other websites with the same domain - eg other websites containing a YouTube iFrame, or pages

---

<sup>80</sup> The `<script>` HTML tag is used to embed a client-side script (JavaScript), ie programs that are processed within the browser. See [w3schools.com, HTML <script> Tag](https://www.w3schools.com/html/html_script_tag.asp).

<sup>81</sup> See section on 'What data does the Facebook Pixel collect?' on the Facebook for Developers pages on GDPR, available [here](#).

<sup>82</sup> Facebook's documentation for developers to implement the Pixel states: "If your website has a Content Security Policy, you should allow JavaScript to load from <https://connect.facebook.net>. Note: The pixel load scripts from two paths: `/en_US/fbevents.js` and `/signals/config/{pixelID}?v={version}`." See [here](#).

on youtube.com. Users may not be aware of iFrames, as it is possible to embed iFrames on a website so that they are completely invisible.

(b) *Widgets* such as social plugins like Twitter's 'Tweet This' or Facebook's 'Like' or 'Share' buttons, can also send data back to the widget provider. For example, Facebook states the data they receive through widgets can include a user ID, the website being visited and other browser information which can be recorded and used for targeted advertising.<sup>83</sup>

74. The embedding of third-party code in websites (pixels and tags) is another example of a commonplace technology, essential to the 'plumbing' of the internet, being used for tracking. Given that the internet relies on HTTP, it would not be practical to abandon this technology in response to the fact that it may be exploited by trackers. In the same way, it would also not be feasible to ban website developers from including code from third parties. Indeed, most software development involves reusing other people's code, and a lot of software is designed to be reused by others.
75. In sum, web tracking technologies are diverse and utilise core features of browsers and the web such as HTTP and JavaScript, which cannot be disabled by users without breaking core web functionality. One common pattern of tracking analytics tools involves blurring the lines of third and first party, whereby sites willingly include third-party tracking code as their own, effectively giving it the permission scope of the first party. This means a user who thinks they are interacting only with the webpage they are on, usually isn't.
76. Some internet governance efforts like the Same Origin Policy have tried to address cross-site scripting, not least because it leads to attacks as well as tracking. However, direct embedding of third-party code without adequate restrictions undermines those efforts. There are efforts to address some of these issues,<sup>84</sup> but they are not in widespread or enforced use. This may suggest there is an opportunity to explore the role of internet governance forums such as the W3C, WHATWG and IETF in enhancing consumer protection online.

### *Mobile apps*

77. Many people who own smartphones carry them wherever they go and rarely share a smartphone with other people. Smartphones are almost always on

---

<sup>83</sup> As [this answer](#) says in the FAQ in Facebook's developer documentation for social plugins.

<sup>84</sup> For example, the [HTTP Feature-Policy header](#) provides a mechanism to allow and deny the use of browser features in its own frame, and in content within any <iframe> elements in the document.

and connected to the internet. They also have a rich variety of identifying data including strong device identifiers like MAIDs, IMEI, IMSI. Furthermore, smartphones are equipped with multiple sensors that can also be exploited for identification purposes.<sup>85</sup> This makes them very personal devices and lucrative targets for trackers. On mobile, trackers may embed themselves as Third Party Libraries (TPLs) or Software Development Kits (SDKs) inside apps (analogous to pixels and tags on webpages),<sup>86</sup> or may passively listen via nearby beacons.

78. On a typical smartphone, there are several providers of user services, each with different roles, relationships, privacy policies and access to user data:
- (a) *Apps* – these can be user-installed (from an app store) or pre-installed (come with the device). Apps may be free to install, or users might have to pay some price. Apps collect and transmit various types and amounts of data about the user, for different purposes. They embed the code of SDKs/TPLs.
  - (b) *TPLs/SDKs* – these are bundles of code written by third parties that apps include into their code to offer extra functionality, for example advertising and analytics services. In current implementations, they inherit all permissions to access user data from the app they are embedded in, for all apps they are included in for a given user’s device.<sup>87</sup> Trackers often take the form of TPLs and SDKs.
  - (c) *App stores* – the marketplaces via which apps are typically installed by users. In principle, users do not need to use app stores to install apps (ie users can sideload apps). In practice an overwhelming proportion of app installs occur via the app store which is bundled with the device OS (App Store for iOS and Google Play Store for Android).<sup>88</sup> App stores provide guidance to app publishers, perform security checks, and can remove apps that do not comply with their guidance. App stores also provide interfaces with app information and user ratings to users who browse.
  - (d) *Operating Systems (OSs)* – Android and iOS are the two main mobile OSs users use today. OSs make the low-level decisions that impact how all apps and other vectors of tracking are managed. They specify the APIs for device-generated data and the permission models for apps that want

---

<sup>85</sup> These include microphone, camera, accelerometer, barometer, GPS, Wi-Fi, gyroscope (can be used for spatio-temporal and biosignatures).

<sup>86</sup> We use the terms TPLs and SDKs interchangeably.

<sup>87</sup> Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., & Gill, C. K. P. (2018). Apps, trackers, privacy, and regulators. In *25th Annual Network and Distributed System Security Symposium, NDSS* (Vol. 2018). Available [here](#).

<sup>88</sup> We note that in some markets, like China, there are other Android App Stores like MyApp (Tencent) which have a larger market share than Google Play Store.

to access user data. More generally, they specify how data is handled generally, and can enforce app store guidelines on a low level.

(e) *Device or original equipment manufacturers (OEMs)* – Android phones are manufactured by various OEMs which are independent from Google, such as Samsung and Huawei. OEMs decide which version of the OS to install<sup>89</sup> and which pre-installed software to add into it. In order to use the Android OS, OEMs must follow compatibility guidelines. iPhones are only manufactured by Apple, which also provides the OS (iOS).

79. In this section we discuss features of the mobile environment that are relevant for trackers. This covers the most common method of tracking on mobile: apps including trackers in the form of TPLs/SDKs, how these TPLs inherit the permissions of their parent app and the consequences. We explain and compare the iOS and Android permission models which apps use to request user data from the device, such as user location, and their shortcomings. We also cover two areas which make tracking more rewarding on mobile compared to desktop, and for which users currently have little control or may not even be aware of: pre-installed apps and sensors.

### *Permissions models*

80. Mobile apps typically run in sandboxes<sup>90</sup> with some limitations on their access to the device’s data. If an app needs data or resources from the device (outside its sandbox), it must declare the appropriate permissions. For example, to access the user’s location data, an app must request the location permission. The permissions model that the OS provides to app developers specifies how to request permissions. Depending on the resource being requested, the OS may grant permission automatically or prompt the user for approval. Since trackers are often TPLs embedded inside apps, they are in turn subject to the same permissions model.

81. Table G.2 gives an overview of the iOS and Android permissions models. These are discussed in turn next.

**Table G.2: Comparison of iOS and Android permissions models**

<i>iOS</i>	<i>Android</i>
<i>Permissions that require user consent</i>	
iOS enforces user consent for any permissions in the Protected Resources list.	Android enforces user consent for any permissions in the Dangerous permission group.
<i>Permission grouping, necessity and minimization</i>	

<sup>89</sup> Android is an open source mobile operating system and there are many versions of Android which OEMs can decide to install. OEMs do not have to install the latest version of Android that Google releases in all their devices.

<sup>90</sup> Application sandboxing is an approach in software development and mobile application management that limits the environments in which certain code can execute.

<i>iOS</i>	<i>Android</i>
<p>One-to-one mapping of permissions to API methods.</p> <p>Apple has implemented a switch for the app developer to state whether the feature is necessary for the app to function or not. This is for hardware compatibility reasons, not data protection.</p>	<p>Groups of permissions are requested. If an app requests permission for one method in a group, all methods in that group are given approval for automatically.</p> <p>Android has implemented a switch for the app developer to state whether the feature is necessary for the app to function or not. This is for hardware compatibility reasons, not data protection.</p>
<i>Inter-app data sharing and permissions</i>	
<p>iOS does not support the sharing of data between different apps and all apps are sandboxed.</p> <p>Apps can be added to a group if all made by a developer. The IDs of the apps must share the same prefix. This enables them to share a directory for storing data.</p>	<p>Android supports sharing data between apps using Content Providers with its permission frameworks.</p> <p>Android also allows apps and OEMs to define their own custom permissions which can enable sharing.</p> <p>Android allows apps to declare the shareUserId attribute if they are signed with the same certificate to access each other's data.</p>
<i>Defaults, timing and design of consent prompts</i>	
<p>Consent prompts at runtime (just in time).</p> <p>Preferences saved if applicable, if not re-prompted.</p> <p>User can change preferences in settings.</p> <p>App developers must check every time in case user has changed settings.</p>	<p>Prompts user for consent at app install and only sometimes at runtime depending on the permission and Android version.</p> <p>No usage description required, but option is available, generic consent screen</p>
<i>TPL/SDK permissions</i>	
<p>TPLs/SDKs automatically inherit the permissions the app is granted. No separate permissions treatment for SDKs.</p>	<p>TPLs/SDKs automatically inherit the permissions the app is granted. No separate permissions treatment for SDKs<sup>91</sup>.</p>

Source: CMA analysis, based on developer documentation for the Android and iOS permission models.

- *Permissions that require user consent*

82. The requirements specify which features require consent to be obtained from the user via a pop-up screen. On Android these are the Dangerous permission group and on iOS these are called the Protected Resources.<sup>92</sup> App developers define which permissions fall under this consent requirement in a configuration file.<sup>93</sup> Note that on Android OEMs (and sometimes apps) may write their own permissions, or rewrite the entire permission system, as unlike iOS, Android is open source, but we discuss here Android's standard API for permissions (we return to the implications of the open source nature of Android in the section on Pre-installed apps).
83. Apple enforces user approval before apps can use Protected Resources. This approval involves a consent screen with some options.

<sup>91</sup> Android's [documentation on content providers](#) states "...components in the provider's application always have full read and write access, regardless of the specified permissions."

<sup>92</sup> A list of Protected Resources on iOS is available [here](#). It includes Bluetooth, calendar, Face ID, location, photos, contacts, files, camera and microphone, homekit, healthkit, music and media, motion (accelerometer), NFC, scripting, system and security configuration permissions, Siri, speech recognition, the user's TV provider account and Wi-Fi.

<sup>93</sup> On Android this file is referred to as the Manifest and on iOS it is the Plist. The documentation for AndroidManifest.xml configuration is available [here](#). Documentation for Apple's Information Property List (Info.plist) is available [here](#).

84. On Android, permissions are grouped according to their ‘protection level’. The developer documentation details which permissions are in which group. The groups are:
- (a) Normal (no user permission required; all apps automatically granted);
  - (b) Dangerous (for personal data, user consent usually required via a consent screen with options);
  - (c) Signature (automatic access given if requester certificate<sup>94</sup> matches applications); and
  - (d) Signature or System (system level access rights).
85. The decisions as to which features should have which protection level may be debatable.
- (a) For example, on Android the `USE_BIOMETRIC` feature is in the Normal permission group.<sup>95</sup> This means application developers do not need to ask a user for consent to obtain their biometric fingerprint.
  - (b) On Android, the `INTERNET` and `ACCESS_NETWORK_STATE` are in the Normal permission group. On iOS, apps also have access to the internet by default and network connectivity is not a Protected Resource<sup>96</sup>. This means application developers do not need user consent to be sending their data on them over the network (away from the device) and reading information about what local networks the device is connected to, which gives coarse location information. Indeed, the Electronic Frontier Foundation have an open request to change the status of the internet permission.<sup>97</sup>
86. On the decision-making process for classifying permissions into protection level groups, Google told us that permissions are classified according to the associated risk that they present to the device user. This depends on the decisions of the maintainers of Android, who are largely Google employees.<sup>98</sup>

---

<sup>94</sup> An app’s certificate identifies the app to Android for authentication and security purposes, in line with [public key infrastructure systems](#). By default Google Play handles this for user-installed apps [with app signing](#). We understand Google must identify the developers by linking their developer account information to these certificate keys it generates for them.

<sup>95</sup> All permissions’ groups are listed next to them as is `USE_BIOMETRIC` in Android’s documentation for the `Manifest.Permission` API which defines its permission model, available [here](#).

<sup>96</sup> The switch to send data from an iOS app over the network is [here](#) and not on their Protected Resource [list](#).

<sup>97</sup> The EFF’s Fix it Already campaign makes several asks of tech companies and is described [here](#).

<sup>98</sup> Here we discuss permissions provided for under the standard Android API (not those defined by OEMs or apps). The standard API depends on code that gets accepted onto the main branch of Android by maintainers, who are historically largely Google employees. The full list of maintainers is available [here](#), and the governance around how they are appointed is [here](#) – note the inertia (only maintainers can nominate maintainers).



- *Permission grouping, necessity and minimization*

87. On Android (unlike on iOS), apps request groups of permissions, not permissions for individual API methods.<sup>99</sup> When an application wants to access an API method, it asks the user for permission for the whole group that API method belongs to. For example, the SMS group includes both READ\_SMS and RECEIVE\_SMS.<sup>100</sup> This may lead to applications having access to more data than they need for their stated purposes. Google confirmed to us that permissions under the Android standard API at Manifest.Permission (not created by the device manufacturer or app developer), currently use a hierarchy of permissions, including individual or group permissions that cluster several similar permissions. If a user approves a permission within a category, all other permissions from that same category are also approved. Android documentation states this is to avoid overwhelming the user with many fine-grained decisions.<sup>101</sup> By contrast, iOS has a one-to-one mapping between API method calls and permissions and groups in their documentation are just decorative (eg Files and Folders).<sup>102</sup>
88. Relatedly, iOS's API for Protected Resources is specified more granularly than on Android. For example, write and read access to photos are separate permissions without a bundled default, whereas Android encourages developers to bundle these and does so by default.<sup>103</sup> Apple is also nudging the user with choice architecture. For example, iOS13 has removed the Always Allow option from the location consent screen.<sup>104</sup>
89. However, neither iOS nor Android explicitly enforce full permission minimisation. This is despite both OSs incorporating a technical switch/flag that could be used to achieve this.<sup>105</sup> This flag was not implemented for data protection reasons, but for hardware compatibility reasons, to facilitate graceful failure in cases where a device does not have the hardware an app relies on. The existence of this flag suggests that if mobile operating systems wanted to encourage or enforce minimisation of permission usage by design it appears technically possible to us.

---

<sup>99</sup> API methods are specific API functions that expose some functionality. For example, READ\_SMS is an API method in the Manifest.Permission API of Android.

<sup>100</sup> SMS stands for Short Messaging Service and is the standard format that text messages take on mobiles. The methods READ\_SMS and RECEIVE\_SMS allow an app with these permissions to read a user's text messages, or to receive and store them itself, respectively. More in the documentation [here](#).

<sup>101</sup> See the documentation on permission groups in Android available [here](#).

<sup>102</sup> See the full list of permissions for Protected Resources on iOS [here](#).

<sup>103</sup> See android.permission 'This attribute is a convenient way of setting a single permission for both reading and writing.' In developer documentation available [here](#)

<sup>104</sup> As this blog from Localytics discusses available [here](#).

<sup>105</sup> This flag/switch is available to app developers when configuring permissions in the Plist file in iOS under the UIRequiredDeviceCapabilities documented [here](#), and in the Manifest file in Android by setting the android:required=[true/false] in the <uses-feature> tag as documented [here](#).



- *Inter-app data sharing and permissions*

90. iOS and Android have different inter-app data sharing mechanisms. Android has methods for apps to expose data to and request data from other apps that do not run in the same process or share an ID<sup>106</sup>, whereas iOS forces apps to be part of the same group if they want to share data.
91. On iOS applications are sandboxed by default meaning they cannot access each other's code and data, and a developer would need to make an App Group<sup>107</sup> for their apps if they want to share data between two or more apps. Similarly, on Android apps are also sandboxed by default, but can group applications by specifying a shared user ID, or signing the app with the same certificate.<sup>108</sup> This grouping effectively just moves the unit of identity from app to developer. A consequence of this design choice is that accountability for the role the app/developer plays in data protection (controller or processor) is still ascertainable, at least in theory.<sup>109</sup> However, on Android, apps can also share data in other ways, between more than one developer's apps.
92. There at least two additional data sharing features Android makes available for apps that do not share the same ID. These are:
  - (a) Content Providers; an app can define a content provider to provide structured access to an app's data for other apps to request access. When defining a content provider, an app can use the permission model to specify access control. By default, data that the content provider makes available is open for reading and writing to all apps. Android encourages app developers to define permissions on their content providers but discourages developers from setting the protection level of these permissions to Dangerous, on the grounds that 'user confirmed permissions...can be confusing for users.'<sup>110</sup>
  - (b) Custom permissions; an app can define specific access requirements to expose some of its data or functionality to other apps. Google confirmed to us that app developers and OEMs can define their own custom permissions. Google informed us that there is no process under which Google reviews the protection level of the permissions that OEMs or app

---

<sup>106</sup> A process is well-defined in computing as an instance of computer program being executed. In this context we can think of it as one app's code being executed in a sandbox.

<sup>107</sup> See the section on Adding an App to an App Group in iOS documentation [here](#).

<sup>108</sup> The documentation for sharedUserId in Android is available [here](#). It notes that apps with the same user ID can access each other's data and, if desired, run in the same process (share a sandbox).

<sup>109</sup> We note that in practice there are several issues with identification due to certificates being self-signed or automatically signed by Google for developers. This is discussed later in the section on pre-installed apps.

<sup>110</sup> Android [documentation](#) states this in two places quoted here, 'In general, we recommend using access controls other than user confirmed permissions where possible because permissions can be confusing for users.' and, 'Each of these poses a significant nontechnical challenge for you as the developer while also confusing your users, which is why we discourages the use of the Dangerous permission level.'

developers define themselves, as Android is open source. We discuss custom permissions and OEMs in more depth in the section on ‘Pre-installed apps’ below.

- *Defaults, timing and design of consent prompts*

93. iOS prompts users for consent when the feature is used (‘just in time’, at runtime), not at install time. iOS’s design gives a range of options such as ‘allow while using app’, ‘allow once’ and ‘don’t allow’. The user’s preference is remembered if applicable and a user can change their preference in the app’s settings or their iOS privacy settings. iOS requires apps to check the authorisation status of a permission every time it accesses a feature as the user may have changed their settings. iOS requires the app developer specify a usage description to go in the consent prompt otherwise access to the permission is prohibited.<sup>111</sup>

94. On Android, the consent prompts are at install time or runtime depending on the Android version and on the permission being requested. Dangerous permissions must prompt the user at runtime in devices running Android 6.0 or higher, and the app has a targetSdkVersion of 23 or higher. If versions are lower than these, all permissions are requested from the user at install time at once. In terms of design/choice architecture, Android runtime prompts have fewer options than iOS, with just ‘deny’ and ‘allow’, although Android recently introduced the one-time option on consent screen for location, camera and microphone.<sup>112</sup> No usage description is required for consent prompts as in iOS, but an option is available.<sup>113</sup>

95. User controls are discussed in a dedicated section later in this appendix.

- *TPL/SDK permissions*

96. On both iOS<sup>114</sup> and Android’s permission models there is no special treatment for embedded TPLs/SDKs. Instead, all SDKs included in an app simply inherit the app’s permissions. This has several implications discussed in the next section.

- *Conclusion on permission models*

97. In sum, a mobile operating system’s permissions model manages crucial decisions regarding default access to user data by apps and their embedded

---

<sup>111</sup> See Apple’s developer documentation on requesting access to protected resources, available [here](#).

<sup>112</sup> As published in Android 11 release notes on permissions, available [here](#).

<sup>113</sup> As the Android documentation specifies [here](#).

<sup>114</sup> On iOS a note warning the developer is in the [documentation](#)

trackers. Notably, mobile OS' and their APIs change frequently as new versions are released, some with significant differences such as Android's introduction of runtime permissions in version 6.0. iOS's permission model tends to have more granularity and less inter-app data sharing, whereas Android uses group permissions, content providers and custom permissions that facilitate inter-app data sharing. Both OSs lack separate permissioning for TPLs/SDKs (they inherit those of the app), and lack any enforcement on necessity of use.

### *Third-party code in first-party mobile apps*

98. TPLs and SDKs are bundles of code that can be included by developers into their apps. TPLs and SDKs offer extra functionality, much like embedded tags/pixels on the web discussed in the earlier section on web browsers. Often this extra functionality is monetisation through personalised advertising and analytics services. TPLs are incorporated into the app by the app developer, much as tags are incorporated by websites, to provide extra functionality for the developer. This section explains how TPLs work, with an example of the widely used Facebook SDK, and some issues with how the permissions model applies to them.
99. The amount of extra functionality TPLs and SDKs offer can be quite substantial. For example, the Facebook SDKs for iOS and Android has multiple components which an app can include individually or together:
  - (a) Analytics – data and trends on app users and how they use the app;
  - (b) Login – a way for users to log into the app with Facebook;
  - (c) Ads – to drive app installs, engagement and use custom audiences for targeting;
  - (d) Share – a way for users to share content they produce on the app on Facebook;
  - (e) App Events – measure effectiveness of ads by measuring app events (user actions);
  - (f) Account Kit (Android only) – app login with a phone number and email (no password); and
  - (g) Graph API – transfer data in and out of Facebook's social graph, query data, post stories, upload photos and other tasks related to Facebook functionalities.

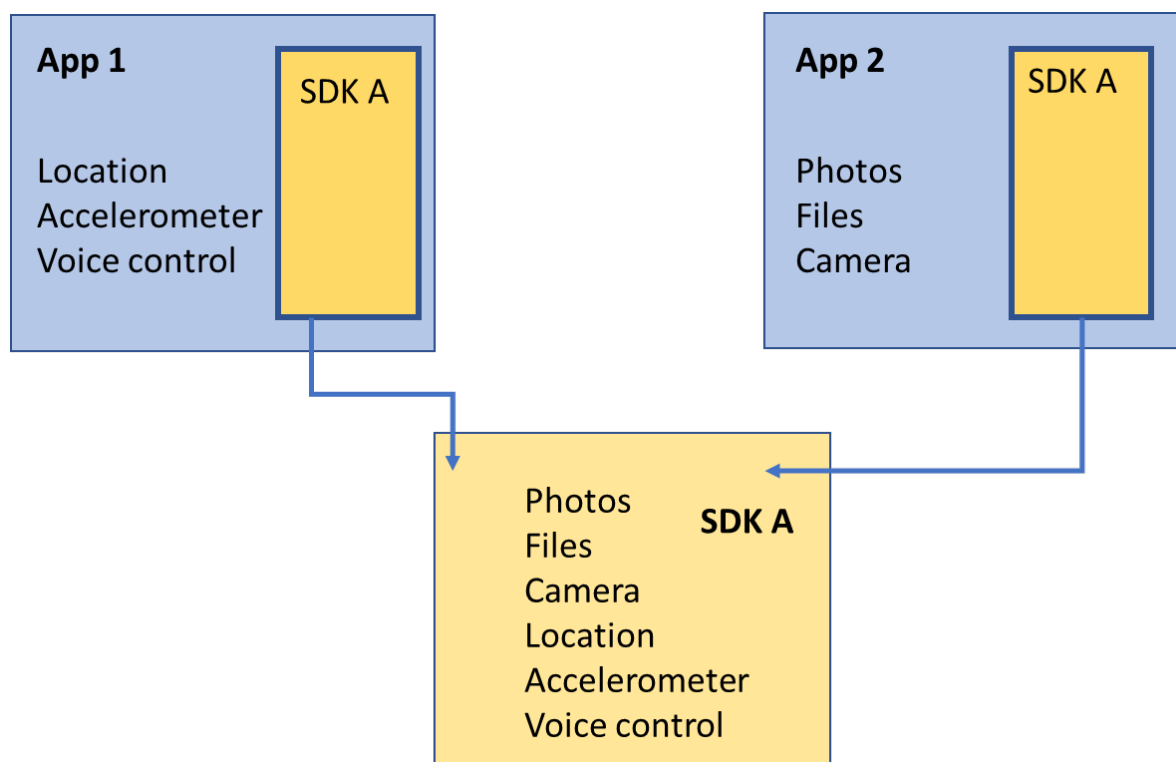
100. This set of functionalities allows app developers to collect significant amounts of data on the user, send it back to Facebook, monetise advertising inventory on the app with Facebook Ads. It also allows apps to integrate Mobile Measurement Parties (which may take the form of a nested SDK).
101. SDKs/TPLs are not visible to the user of the app. This may lead the user to think that they are interacting only with the first party (ie just the app).
102. SDKs/TPLs inherit permissions to access user data from the app in which they are embedded. This means TPLs automatically have access to the same user data the app does. This is significant for two reasons:
  - (a) Firstly, a user may reasonably believe that when they consent to an app using a given piece of personal data, such as location data, they are only agreeing that the data will be used by the app itself. However, in practice they would also granting access to all TPLs the app embeds, and neither the app nor the OS informs the user of this or gives them any options.<sup>115</sup>
  - (b) The second issue with TPLs inheriting the app's permissions can arise when a TPL is embedded in multiple apps on a user's device, which is often the case. This effectively gives the TPL the union of permissions across all apps they are embedded in, which can cover a broad set of permissions. This is illustrated in Figure G.5: the list of permissions in the large yellow box shows what access rights SDK A inherited by virtue of being included in App 1 and App 2. SDKs only need to be integrated in a few apps in order to have access to most of a user's data. Note also that TPLs can be embedded in other TPLs, as with Facebook's Mobile Measurement Partners.<sup>116</sup> This means a user's personal data may be processed multiple degrees away from the original party they engaged with. This is also true for websites that embed JavaScript, discussed earlier.

---

<sup>115</sup> European Union Agency for Network and Information Security - ENISA (2017). Privacy and data protection in mobile applications. A study on the app development ecosystem and the technical implementation of GDPR. Available for download [here](#).

<sup>116</sup> Their [FAQs on the Facebook SDK](#) say mobile measurement partners integrate directly with Facebook. We interpret this to mean it is integrated in their SDK.

**Figure G.5: Illustration of how SDKs inherit the union of permissions of all apps they are embedded in.**



Source: CMA.

103. A 2018 study called ‘Apps, Trackers, Privacy and Regulators’ characterised the global extent of activity of advertising and analytics TPLs in Android phones.<sup>117</sup> These researchers found that the average mobile app connects to 11 different third-party domains, of which six are used for tracking. They also found that the most common value harvested by advertising-related trackers is the AAID, and that in 34% of the cases the AAID was collected in conjunction with another persistent unique device identifier (such as the IMEI), against Android’s Developer Guidelines.<sup>118</sup>
104. A 2016 study also showed that 70% of apps dedicate at least 10% of their traffic to advertising trackers, and 7% dedicate 90% of traffic to them – suggesting many apps would operate offline and use less mobile data from the user if not for advertising trackers.<sup>119</sup> The extra traffic use suggests that, despite the benefits of extra functionality they offer, TPLs come at a latency

<sup>117</sup> Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., & Gill, C. K. P. (2018). Apps, trackers, privacy, and regulators. In *25th Annual Network and Distributed System Security Symposium, NDSS* (Vol. 2018). Available [here](#).

<sup>118</sup> In [its developer guidelines](#), Android state, ‘The advertising identifier must not be connected to personally-identifiable information or associated with any persistent device identifier (for example: SSAID, MAC address, IMEI, etc.) without explicit consent of the user.’

<sup>119</sup> Vallina-Rodriguez, N., Sundaresan, S., Razaghpanah, A., Nithyanand, R., Allman, M., Kreibich, C., & Gill, P. (2016). Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem. arXiv preprint arXiv:1609.07190. Available [here](#).

cost to the app and therefore a cost to the user in terms of experience as well as mobile data cost.

105. In sum, TPLs/SDKs are used to embed third-party code in mobile apps. These TPLs/SDKs offer a lot of additional functionality, including advertising and tracking services. On average, a dozen of these TPLs/SDKs are present in an app. On Android, a TPL's permissions to access user data are inherited from the app it is included in. If TPLs are embedded in even just a handful of apps on a user's device, they may have access to a very broad range of the user's data.

### *Pre-installed app ecosystem*

106. So far, we have discussed apps without distinguishing between user-installed and pre-installed apps. In this section we highlight some unique aspects of pre-installed apps. Pre-installed apps are apps that are installed by the OEM on a device prior to its use, rather than being installed by the user. They are sometimes called 'bloatware',<sup>120</sup> because they are often unwanted by the user and take up device memory.
107. iOS's pre-installed app ecosystem is far simpler than Android's. This is a consequence of Apple's integrated model, where the devices are designed and manufactured by the same entity. All pre-installed apps on iOS are Apple-owned.<sup>121</sup> By contrast, many OEMs run the open source Android OS on their devices. It is relatively difficult to obtain a list of all the pre-installed apps on Android devices from all the OEMs that run Android.<sup>122</sup> In terms of user control, pre-installed apps can be removed by the user on iOS,<sup>123</sup> whereas on Android they cannot.<sup>124</sup> For this reason, this section focuses exclusively on the Android pre-installed app ecosystem.
108. One key difference between user-installed and pre-installed apps is the low user awareness of the existence of pre-installed apps. A recent study by Gamba et al. on pre-installed apps in Android is the most comprehensive investigation of this topic that we are aware of.<sup>125</sup> The authors scanned 2,700 users' devices, spanning 1,700 device models running Android from 214 OEMs. They found that only 9% of these pre-installed apps were also

---

<sup>120</sup> For example, in this [article](#) on Wired.

<sup>121</sup> A full list of all pre-installed apps for the latest version of iOS by Apple is available [here](#).

<sup>122</sup> Google told us that they did not have access to a list of all the pre-installed apps on Android devices. It seems all OEMs would need to be asked for this information, and they also may contract this out to third parties as with [ironSource's out-of-the-box experience](#).

<sup>123</sup> As Apple describe [here](#).

<sup>124</sup> There is no easy way to uninstall all pre-installed apps on Android devices. Often they can be disabled but not uninstalled without rooting the phone and using highly technical super user methods or installing third party software which might pose additional risks, as described [here](#), that most consumers will be unable to do and may pose risks.

<sup>125</sup> Gamba et al. (2019) *An Analysis of Pre-installed Android Software*. Available [here](#)

available to download in the Google Play store. The study notes, ‘the low presence of pre-installed apps in the Play Store suggests that this type of software may have escaped any scrutiny by the research community’.

109. Gamba et al. found that the median pre-installed app requests access to three Dangerous permissions. For user-installed apps, this class of Dangerous permissions requires user consent to be given just in time (at runtime), as discussed in the section above on permissions models. It is unclear whether the pre-installed apps requesting Dangerous permissions are doing so under the same explicit runtime consent mechanisms. Indeed, Gamba et al. have expressed concern that it is unclear whether users are given any opportunity to consent to the practices of pre-installed apps. To investigate this, the authors acquired six popular Android devices; they found that, upon the first initialisation of each device, three presented only the Android terms of service, whilst the other three showed a privacy policy mentioning personal data is collected for added value services. In all cases, users have no choice but to accept these terms, otherwise the device will not boot and be therefore unusable. This suggests that, in some cases, the user may have little choice but accept pre-installed apps. Once installed, pre-installed apps often cannot be removed by the user.<sup>126</sup>
110. OEMs or pre-installed apps may define their own custom permissions on Android.<sup>127</sup> Apps that define custom permissions can share their resources and capabilities with other apps. Although both user-installed and pre-installed apps can define custom permissions, the concern is more acute for pre-installed apps: there is evidence that they may allow other apps to obtain access to privileged system resources and sensitive data in a way that circumvents the Android permission model, which we discuss presently.
111. Once written, custom permissions can be used by other actors on the hardware/firmware level, since they are often defined in core Android modules (Figure G.6), circumventing a lot of the protections for user data that Android has implemented. Some examples that Gamba et al. document include:
  - (a) Custom permissions are used to implement proprietary VPN<sup>128</sup> solutions. For example, Samsung and Meizu do this<sup>125</sup>. These VPNs allow

---

<sup>126</sup> This is in Gamba et al. cited above. Also Lifewire discuss this [here](#), as does [this top answer](#) on Android StackExchange, uninstalling cannot be done without rooting the phone – a highly technical endeavour with some risks to the device’s integrity and functioning. Some developers have created applications to do this on behalf of a user, as described in this [article](#) on devsjournal, but they are convoluted, require technical skill, and may introduce new risks (they all require privileged access themselves). Sometimes Android allows a user to disable but not uninstall pre-installed apps.

<sup>127</sup> We confirmed this with Google. Custom permissions can be written by both user-installed and pre-installed apps and it is one way they can make data available to other apps. The documentation is [available here](#).

<sup>128</sup> [Virtual private network](#) – extends a private network over a public network enabling communication as if both computers were on a shared private network.

circumventing of Android’s sandboxing and let an app monitor a user’s traffic device wide.**Error! Bookmark not defined.**

- (b) Baidu’s geo-location permission is exposed by pre-installed apps including core Android ones.**Error! Bookmark not defined.Error! Bookmark not defined.** This allows circumvention of Android’s location permission, which is classified as Dangerous and requires user consent.
- (c) Gamba et al. found a pre-installed app signed by<sup>129</sup> Vodafone (Greece) in Samsung devices exposes a custom permission associated with Exus, a credit risk and banking company. This suggests that pre-installed apps may be a relevant source for data brokers.

**Figure G.6: Summary of custom permissions by provider category and their presence in selected sensitive core Android modules.**

	Custom	Providers							
	permissions	Vendor	Third-party	MNO	Chipset	AV / Security	Ind. Alliance	Browser	Other
<b>Total</b>	4,845 (108)	3,760 (37)	192 (34)	195 (15)	67 (63)	46 (13)	29 (44)	7 (6)	549 (75)
<b>Android Modules</b>									
android	494 (21)	410 (9)	—	12 (2)	4 (13)	—	6 (7)	—	62 (17)
com.android.systemui	90 (15)	67 (11)	1 (2)	—	—	—	—	—	22 (8)
com.android.settings	87 (16)	63 (12)	—	1 (1)	—	—	—	—	23 (8)
com.android.phone	84 (14)	56 (9)	—	5 (2)	3 (5)	—	—	—	20 (10)
com.android.mms	59 (11)	35 (10)	—	1 (2)	—	—	1 (1)	—	22 (8)
com.android.contacts	40 (7)	32 (3)	—	—	—	—	—	—	8 (5)
com.android.email	33 (10)	18 (4)	—	—	—	—	—	—	15 (17)

Source: Gamba et al. (2019) *An Analysis of Pre-installed Android Software*. Available [here](#).

Note: Bracketed values are the number of OEMs in which these custom permissions were found.

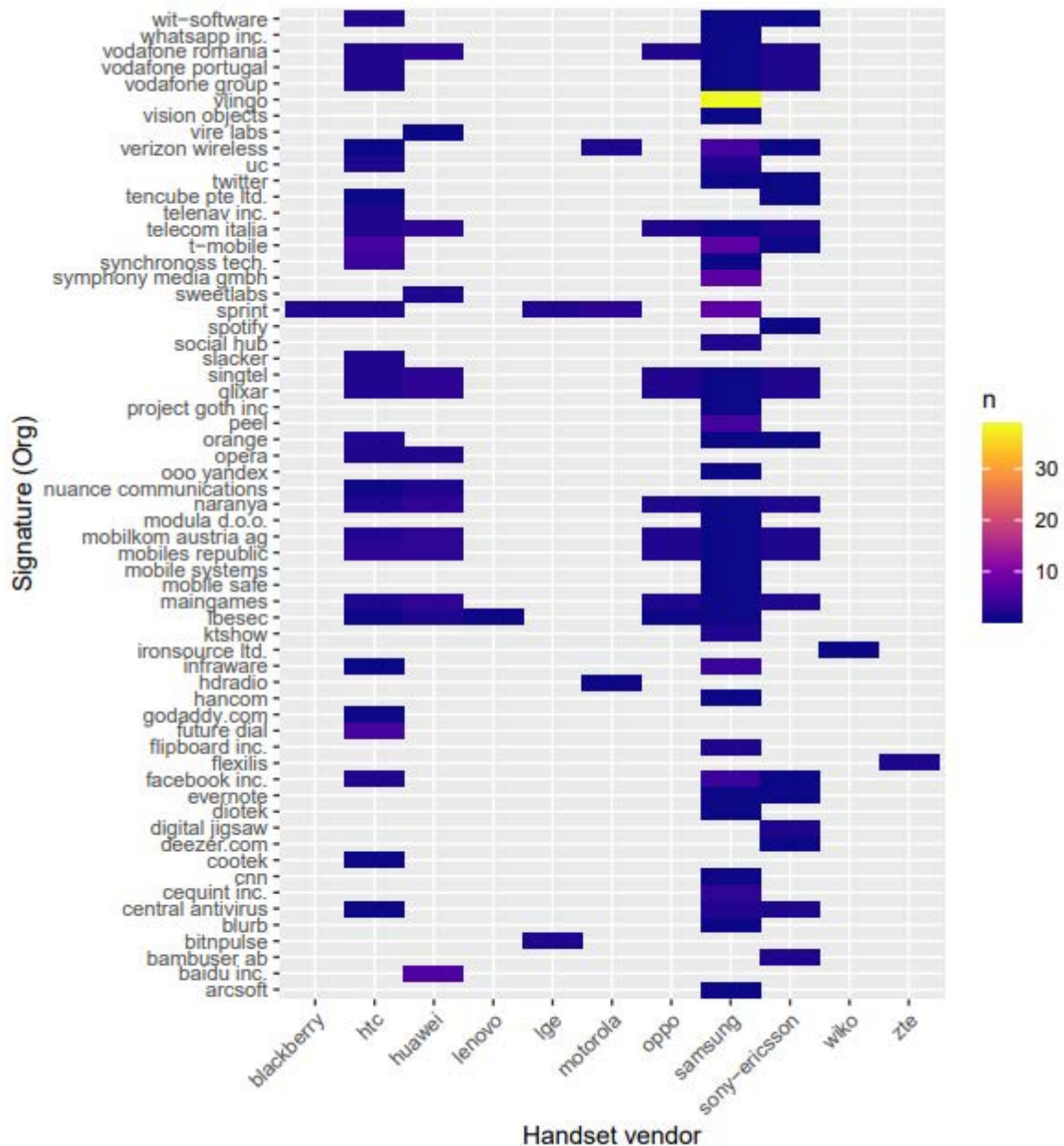
112. It’s useful to consider estimates of how widespread custom permissions are. Gamba et al. identified 1,795 APKs<sup>130</sup> across 108 vendors defining 4,845 custom permissions, excluding Android/Google made ones. Three OEMs accounted for 68% of custom permissions: Samsung (41%), Huawei (20%) and Sony (7%). These are used by apps users can install from the Play store, as shown in Figure G.7 below. But custom permissions are also defined by parties with whom the user has had no interaction (unlike the OEM, from whom they purchased the device). For example, the authors found six pre-installed apps from Facebook defining 18 custom permissions and three of these apps were unavailable on the Play store.

<sup>129</sup> [Certificate signing](#) is a process all apps do to digitally sign their public key which is indented to identify them for secure communication purposes

<sup>130</sup> APK (Android package application) is the file format for Android for distribution and installation of Android apps.



Figure G.7: Apps accessing OEMs custom permissions.



Source: Gamba et al. (2019) *An Analysis of Pre-installed Android Software*. Available [here](#).

113. As with the user-installed app ecosystem, pre-installed apps can also include TPLs, many of which are advertising and analytics trackers. The kinds of permissions that TPLs access can be seen in Figure G.8. Many TPLs can read logs, and mount and unmount file systems. Many also have the WRITE\_SECURE\_SETTINGS permission,<sup>131</sup> especially social media services, despite Android documentation explicitly stating that these permissions are ‘Not for use by third-party applications’.<sup>132</sup> Figure G.8 shows

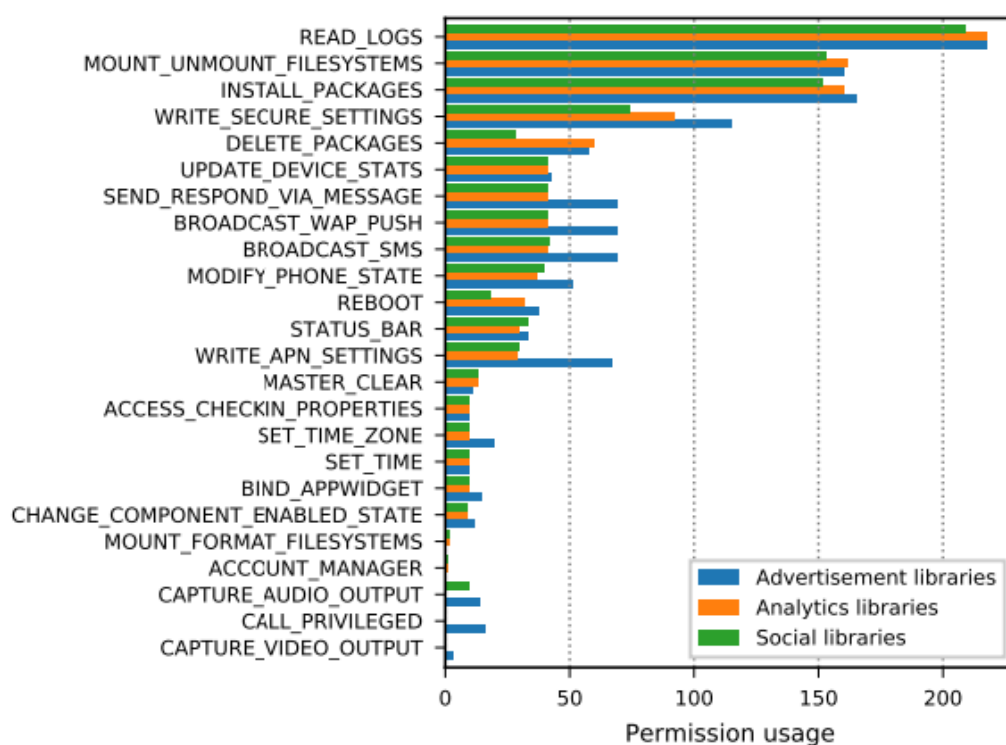
<sup>131</sup> This allows an application to write secure system settings, which normally they can only read. Doing this can override any preferences a user has set in their settings. According to our understanding, the full list of what can be overridden is in the Settings.Secure class [here](#).

<sup>132</sup> In Android’s Manifest.Permission API [here](#), which defines all permissions except those that are custom defined.

advertising TPLs disproportionately use WRITE\_APN\_SETTINGS, which is used to write network settings – a very privileged action which can reconfigure network security and if/how a mobile device should connect to some private customer network. Obtaining such low-level permissions is easier for trackers that embed themselves as TPLs in pre-installed apps than in user-installed apps because, as discussed above:

- (a) Pre-installed apps may not require the user’s consent for every Dangerous permission; as discussed above Gamba et al. found often they were bundled into a single small print notice, which the user must accept in order to use the device at all.
- (b) Pre-installed apps have privileged access to system resources, especially if signed with the platform certificate (discussed soon), as illustrated in the examples of custom permissions above.

**Figure G.8: Permission usage by pre-installed apps embedding TPLs**



Source: Gamba et al. (2019) *An Analysis of Pre-installed Android Software*. Available [here](#).

- 114. The decision as to which pre-installed apps a device is equipped with is made by OEMs in agreement with the developers of pre-installed apps, who are third parties from the user’s point of view. Google told us that OEMs do not need to seek approval from Google to pre-install apps and do not need to provide a list of pre-installed apps present on their device by default.
- 115. We asked Samsung, a popular OEM, about their pre-installed apps and agreements with Google. Samsung told us they entered into a Mobile

Application Distribution Agreement (MADA) with Google. We understand that OEMs enter into MADAs with Google which grants them license to use Google Mobile Services on their devices.<sup>133</sup> This requires passing the Compatibility Test Suite (CTS), an automated self-testing suite OEMs run to check compatibility and compliance with Android as defined by the Compatibility Definition Document.<sup>134</sup> This would make the CTS a natural home for checking permission model circumvention by pre-installed apps. However, as far as we know, currently the CTS tests only for safety and compatibility, not for security<sup>135</sup> or protecting the user's data<sup>136</sup>.

116. Samsung could not confirm which TPLs/SDKs are included on apps they pre-install (SDKs are the form trackers often take), they said that each app developer has this information. TPLs/SDKs tend to inherit the identity of the app as well as its permissions, even though they are developed by a third party developer, not the app developer.
117. More generally, the problem of identifying actors in the Android ecosystem is particularly hard because Android is open source and when OEMs customise Android by adding pre-installed apps, they may sign all the software together with one certificate (often known as the 'platform certificate'). As discussed above in the section on 'Inter-app data sharing and permissions', when two apps are signed by the same certificate (or share a user ID<sup>137</sup>) they automatically get access to each other's data and permissions. This is analogous to websites being exempt from the Same Origin Policy on the web if they were signed with the same SSL certificate, and is included under our definition of tracking.<sup>138</sup>
118. More generally, apps (both user and pre-installed) can self-sign certificates. This contrasts with on the web, where self-signed certificates are normally

---

<sup>133</sup> Google Mobile Services include the Google Play Store, Search, Chrome and YouTube. The Android OS itself is open source.

<sup>134</sup> Compatibility and compliance for OEMs who want to run Android on their devices is defined by the Android Compatibility Definition Document [available here](#).

<sup>135</sup> Gamba et al. noted that pre-installed apps are insecure and rarely get updated, making them more vulnerable to malicious attacks. For example, they found that 74% of the non-public apps (pre-installed apps that are not in the app store) are never updated, and 41% remain unpatched for five or more years – meaning a user may be at risk from this for as long as they typically own their phone.

<sup>136</sup> Google told us CTS is used to enforce compliance with the 'Pre-grant Permission Policy' which requires OEMs to include user prompts in accordance with the Android runtime permission model on devices running Android v.6.0 or higher. Google told us exceptions are granted for apps required for core functionality or to set-up the device, or to enable certain emergency services on the app. Google instructs OEMs how to do this in the Android developer documentation [here](#). We note that the OEM can simply declare any app to be a core service or default handler on trust.

<sup>137</sup> The sharedUserID attribute as documented [here](#). Note Android is deprecating this in version 10, API level 29, released on September 2019; of course, apps may be targeting earlier versions of Android.

<sup>138</sup> Our definition of tracking includes cases where the owner of the property is the same but the properties themselves present as distinct to the user (for example, youtube.com and google.com). This is because a user cannot expect to know who owns which properties, and what the internal data sharing practices of the company are.

flagged as potentially insecure by the browser. On the web, typically signing is done via one of many trusted Certificate Authorities.<sup>139</sup> On mobile, to our knowledge, there are no trusted certificate authorities and OSs do not enforce this (unlike browsers).

119. The certificate provenance issues make identities and partnerships difficult to disentangle for researchers (and for regulators). For example, Gamba et al. found apps with certificates signed by Sprint<sup>140</sup> resembled some apps of Facebook's requesting Flurry-related<sup>141</sup> permissions. In this way up to three mobile actors (an MNO, and two apps/SDKs) are sharing a user's data without their awareness. Disentangling these identities was a challenge for Gamba et al. due to self-signing (the authors noted many certificates simply signed with 'Debug').
120. Gamba et al. conclude their paper with some recommendations for regulators and policymakers. These include first and foremost the introduction and use of certificates that are signed by globally trusted certificate authorities, or repository which provides certificate details publicly for accountability. They call also for more accessible documentation and consent forms. They note that, 'similar to the manner in which open-source components of Android require any modified version of the code to be made publicly available, Android devices can be required to document the specific set of apps that have pre-installed, along with their purpose and the entity responsible for each piece of software...'
121. In sum, pre-installed apps in Android present a series of opportunities for user tracking. Users cannot choose whether to have these apps installed and may not be aware of the existence of pre-installed apps (only 9% of pre-installed app are available in the Play Store). Users cannot remove many pre-installed apps on Android, only disable them. Furthermore, due to the open source nature of Android, there are cases where pre-installed apps circumvent the Android permissions model, implementing custom permissions or being signed with the same certificate, being automatically granted access to Dangerous and Privileged permissions. All of this makes the ecosystem of pre-installed apps on Android one of the least understood and potentially most important areas for understanding how trackers obtain user data in the mobile sphere.
122. We note that successive versions of Android introduce changes frequently (including to the permissions model) and each OEM is free to modify and

---

<sup>139</sup> Certificate Authorities (CAs) issue digital certificates, small files that contain identity credentials to help websites and devices identify themselves online, read more [here](#) on GlobalSign's blog.

<sup>140</sup> Sprint is an Mobile Network Operator (MNO), their website is [here](#).

<sup>141</sup> Flurry is a mobile app analytics company, their website is [here](#).

implement Android differently and to retain old versions of Android on their devices, it is difficult to draw general conclusions and the precise details of what we discuss in this section may not apply in every case. However, the issues we have set out (on pre-installed apps, custom permissions, and certification provenance issues) deserves further attention.

### *Mobile sensors*

123. Mobile devices such as smartphones generate personal data well beyond emails, pictures, contacts and chat messages. Device-generated data that can support tracking: includes identifiers such as IMEI, IMSI, MAIDs, logs and other metadata that can be used in fingerprinting, and sensor-generated data. We explore the latter in this section.
124. There are various sensors embedded in mobile devices that collect user data that is potentially valuable for tracking.<sup>142</sup> These include: (i) time and location; (ii) thermometer; (iii) barometer; (iv) accelerometer and gyroscope.
125. Relatively small samples obtained from these sensors can uniquely identify a user. For example, one study<sup>143</sup> found that just four spatio-temporal data points are enough to identify 95% of people. Similarly, motion sensor data (accelerometer and gyroscope) for just five human steps is needed for a user's unique gait to be fingerprinted<sup>144</sup>. Battery capacity was also found to be an effective data source for unique device fingerprinting.<sup>145</sup> Other seemingly trivial data sources can be used to fingerprint a mobile device, even excluding sensor data – such as list of apps installed.<sup>146</sup>
126. Many smartphone sensors can be sources of location data:
  - (a) Mobile Network Operators (MNOs) can use triangulation to locate a device identified from their unique SIM identifier, the IMSI. Using the signal strength that different towers observe for a unique device, they calculate location to an accuracy equivalent to around a city block.
  - (b) Wi-Fi and Bluetooth antennas operate on short range waves and transmit the device's MAC address identifier. These are very commonly used for

---

<sup>142</sup> Personal computers (such as laptops) also include sensors, but they are less likely to be used “on the go”.

<sup>143</sup> De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 1376.

<sup>144</sup> Gadaleta, M., & Rossi, M. (2018). Idnet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition*, 74, 25-37.

<sup>145</sup> L. Olejnik, G. Acar, C. Castelluccia and C. Diaz, “The leaking battery: A privacy analysis of the HTML5 Battery Status API,” 2015.

<sup>146</sup> One [study](#) achieved 97% accuracy in fingerprinting a device with simple models using attributes that seem non personal and are easily accessible via API without need for user consent (in Android's Normal permission group) including device names, language settings, lists of apps installed, most played songs.

location-based and indoor advertising: among other tasks, they can be used to infer whether a shopper has entered a building.<sup>147</sup>

(c) Geolocation based on Global Positioning System (GPS) works using satellites to triangulate a device's location into longitude and latitude pairs. These are calculated by the OS, but can be exposed to apps through the operating system's permissions model

(d) RFID and NFC are known as proximity sensors, and operate on smaller distances than GPS. They are typically used for 'out of home marketing'.<sup>148</sup>

127. It is worth highlighting that it is the *stream* of location data (that is, sequences of data points over time) that enables rich profiling of a consumer, including where they work, live, key places of interest, down to interactions with other individuals. Furthermore, aggregating data for multiple connected users into a social graph can give more insight<sup>149</sup> on target audiences and recommendations for personalised advertising<sup>150</sup>

128. Overall, smartphones provide a rich ecosystem for trackers due to a number of features: their personal nature, being operational at most times, often connected to the internet, equipped with easily accessible device IDs such as IMSIs and MAIDs, and the availability of rich sensor data for fingerprinting. Most trackers on mobile take the form of TPLs/SDKs embedded in user-installed and pre-installed apps. Apps are subject to the permission models of the OS to access user and device data, but TPLs simply inherit permissions from the app they are nested into. This can enable TPLs to access a range of user data, even when embedded in just a handful of apps. Users are mostly unaware of the presence of TPLs, just as they are unaware of the pre-installed apps. Pre-installed apps are less scrutinised than user-installed ones, and multiple issues have been found in Android by researchers. In terms of actors, mobile operating systems (OSs), device manufacturers (OEMs) and mobile networking operators (MNOs) play the largest role in stewarding user mobile device data (although apps have a role too, especially in including TPLs).

---

<sup>147</sup> As described in this blog post on proximity marketing by beaconstac available [here](#).

<sup>148</sup> For example, Estimote's SDK allows an app to send personalised push notifications to a user upon entering a shop.

<sup>149</sup> Databricks discuss building a location based social graph [here](#).

<sup>150</sup> Microsoft compile a list of papers that use the social graph for intelligently recommending personalised advertising [here](#). the 'Selected Representative Research' tab is recommended.



## User control and tracking

129. Tracking is technical in nature, and most consumers will be unaware they are being tracked. Those that are aware may feel limited in what they can do to prevent being tracked, and might consider being tracked as an unavoidable feature of the service they are using. Although some tracker prevention innovation is happening, consumers still largely lack control and choice on what data is collected on them, or how it is shared, linked and used. Even for sophisticated users, avoiding being tracked is a difficult endeavour; information asymmetries for the average consumer are likely to be extreme.
130. This section covers the extent to which users have control over how their data is collected and used by advertising and analytics tracking services, with a range of examples building on the section on tracking technologies above. We highlight the role of major browsers and mobile operating systems that, in practice, make many decisions for users and act as their agent.
131. The tracking ecosystem presents challenges for user choice and control due to information asymmetries. User awareness of tracking technologies and their implications is low, as discussed in Appendix L. Some examples of information asymmetries in tracking are:
  - (a) Upon visiting a website or app, a user may reasonably think they are interacting only with that website or app. However, pixels/tags and TPLs/SDKs are often embedded in websites and apps, collecting and sending data to third parties. There are often no visual clues to the user as to the presence of most TPLs, except for social media plugin buttons and widgets. Typically, these TPLs will inherit the permissions of the site or app they are embedded in, which allows for cross-app tracking as discussed earlier.
  - (b) On Android, users may be unaware of the many pre-installed apps that their device comes with. Even if they were to become aware of them and decide not to use them, they would only be allowed to disable them, not uninstall them. Furthermore, most pre-installed apps are not designed for users, as only 9% of them appear in the app store. See the section on pre-installed apps earlier in the appendix for more details.
  - (c) Users may be unaware that data brokers exist, and that they collect user data and sell it to other parties. Additionally, some of these data have been shown to come from trusted personal devices<sup>151</sup>.

---

<sup>151</sup> Such as in the cases uncovered by Motherboard where MNOs such as AT&T, T-Mobile and Sprint were selling real-time location data to data brokers, with some of this making its way to bounty hunters. Available [here](#).



132. The tracking ecosystem, especially on Android, contributes to what we call in Chapter 4 a ‘take it or leave it’ dilemma. We discuss this issue beyond tracking in the section on the impact of the consumer choice remedy in Appendix X. However, here are some examples particular to tracking that this appendix has discussed:

- (a) Pre-installed apps must be accepted in order to boot an Android device;
- (b) The AAID is a unique, persistent and available identifier to all apps on Android. All Android devices come with AAIDs, which cannot be completely disabled. Recently, NOYB<sup>152</sup> filed a legal case against Google that processing the AAID is in breach of the GDPR. They allege that:<sup>153</sup>
  - (i) NOYB made a subject access request under Article 15 of the GDPR. In its response, Google did not provide specific, complete and updated information on the processing of the AAID, such as its recipients, sources, legal basis or retention periods, and instead referred to its privacy policies.
  - (ii) Google lacks an initial legal basis (within the meaning of Article 6 of GDPR) to generate the AAID. Google relies on consent to the Google privacy policy, which is neither informed, specific (as the data subject has to agree to processing by all Google services within a single step), nor free (as user could not use their phone without agreeing), and is thus invalid. If Google were relying on legitimate interest, then NOYB further disputes that the interests of Google and other companies in tracking users for advertising overrides the privacy interests.
  - (iii) Google’s proposed solution (to reset the AAID) does not stop the processing of personal data, because resetting is not the same as turning off and the new ID can be easily linked to the old one.<sup>154</sup> Google states that it does not link successive AAIDs of a specific user who has reset their AAID, but note that this does not preclude apps<sup>155</sup>, OEMs or MNOs from doing so. The way that Google has implemented the AAID makes it impossible to prevent the processing of personal data within Android devices. This is in contrast to Apple,

---

<sup>152</sup> NOYB stand for “none of your business” and are a not-for-profit based in Austria that aim to help the enforcement of GDPR on a European level. More information is available on their website [here](#).

<sup>153</sup> The full NOYB complaint is [available here](#).

<sup>154</sup> The NOYB complaint discusses the technical ease of linking of identifiers as we have earlier in this appendix. We discuss this for the controls on the AAID specifically in the next section.

<sup>155</sup> However, in their Developer Guidelines (available [here](#)) Google do ask app developers not to link the new AAID of users who have recently reset AAID to these users’ old AAID. NOYB argue that the fact that Google asks developers not to do so proves that Google has no technical ability to stop the processing of data and tracking using AAIDs.

which allows the IDFA on iOS to be effectively disabled by setting it to a string of zeros.

133. The lack of user awareness and information asymmetries show the importance of fairness in choice architecture, discussed in Appendix Y (on Fairness by Design). In the tracking ecosystem, a fair choice architecture is particularly challenging due to the nature of technology enabling tracking in a way that cannot be easily turned off. For example,
- (a) As discussed earlier, fingerprinting exploits the combination of multiple identifiers to identify individuals. Almost any piece of information can be used to help identify individuals, no matter how impersonal it appears in isolation.<sup>156</sup>
  - (b) The web itself is enabled by HTTP communication. However, the significant data storage ability in this protocol can be exploited by trackers, for example, in request headers and in parameters. It is not easy to change this design because some metadata is needed in order to verify HTTP requests.
  - (c) The use of TPLs/SDKs and embedded JS libraries is commonplace and most developers rely on reusing other people's code as a routine part of web and app development<sup>157</sup> and restricting their use altogether would not be a viable proposition.
134. The above suggests that in most cases users don't have an option to straightforwardly 'turn off all tracking'. More work is warranted to understand the trade-offs in this area. This is something to bear in mind when designing any interventions. Some initiatives to limit tracking have gone to great lengths and are still improving (for example, Apple's Intelligent Tracking Prevention on Safari, which we discuss shortly).

### ***Controls over tracking***

135. Despite the invisibility and lack of effective control a user might have over being tracked and their data being used without their knowledge or consent, there are some specific examples of controls that platforms do or do not provide. This may include to users' ability to:

- (a) turn off data collection;

---

<sup>156</sup> The fact that there are many pieces of information/vectors for a tracker/attacker, and only a few are needed for successful identification/exploit to happen has parallels with the problem of wide attack surfaces in cybersecurity. Attack surface is discussed comprehensively by OWASP [here](#). More generally, approaches to protect users from trackers can learn from approaches and methods in cybersecurity.

<sup>157</sup> The government's technology service manual on software dependencies supports this view, available [here](#).

- (b) clear or remove data already collected on them; and
  - (c) remove an unwanted feature or app.
- 136. In practical terms, once a decision to provide users with a certain control is deemed desirable, two main issues remain:
  - (a) What should the design/user journey be like to maximise engagement?
  - (b) Where in the software layers should the controls be implemented?
- 137. We do not address the first question here, which is discussed in Appendix Y on Fairness by Design. With respect to the second question, we consider that it is generally easier if such options are implemented by lower-level software platforms<sup>158</sup>, both for consistency and for ease of implementation. For example:
  - (a) If every app had to implement its own notices for gaining user consent to use Dangerous permissions such as location, the total effort involved would have been much higher than if it were implemented by the operating system.
  - (b) If changes were implemented by browsers, it would place less burden on websites to adapt their practices themselves, bypassing coordination and enforcement issues.<sup>159</sup>
- 138. As the underlying software that many websites and apps are built upon, OSs and browsers would be well-positioned to take the role of stewards for user control, enforcing defaults that protect the user and enhance their experience. However, we note changes to lower-level software, eg OSs and browsers, can influence the trajectory of the entire industry.

## *Mobile*

- 139. An example of controls given in the mobile ecosystem is the permissions models discussed in earlier. Mobile OSs use permissions to regulate apps' access to personal data on from the device. Besides protecting system resources, this is designed to give the user control over personal data that might otherwise be accessed by apps without their knowledge. Typically, some of the data protected by the permissions models fall under the definition

---

<sup>158</sup> By lower-level here, we mean the software hierarchy where at the bottom is a the device's hardware, followed by system software such as operating systems, followed by particular 'entry point' platforms such as browsers or app stores, and finally websites, apps and TPLs/SDKs/tags/pixels.

<sup>159</sup> This type of coordination and enforcement issues are illustrated by the failure of the Do Not Track (DNT) initiative, which [has recently become unsupported](#) by Safari. It's plausible that DNT failed to take off because of the work required from websites.

of personal data as described in GDPR, such as location data, but it is unclear how the protection level of a given category of data is currently determined by mobile OSs.

140. Controls on MAIDs<sup>160</sup> vary by mobile OS. On iOS the IDFA can be disabled when the user selects ‘Limit Ad Tracking’ in their settings. By contrast, the ‘Opt out of Ads Personalization’ setting in Android does not remove the ability for apps to access the AAID, but instead requires apps to be responsible for actively seeking out the user’s preference about whether their data is used for targeting ads and honouring it.<sup>161</sup> The user is able to reset the AAID however, and a new one is generated which may be linked back to the old one by apps.<sup>162</sup> We further note that Android app developers are, in any event, permitted to track users via the AAID to perform frequency capping and conversion tracking, and that unlike for iOS there is no OS-level setting for a user to opt out of tracking for advertising entirely.<sup>163</sup>
141. On Android users are unable to remove pre-installed apps, while this is possible on iOS. Individual pre-installed apps can be disabled, but not uninstalled, and not all at once. Instead, rooting the phone or similarly highly technical measures must be taken to remove pre-installed apps from an Android mobile.<sup>164</sup>

## Browsers

142. Several browsers have taken active measures to counter tracking and protect users, in their role as a user agent. In this section, we look at browser-level

---

<sup>160</sup> As discussed earlier, the MAIDs are unique and persistent device identifiers, available to all apps, which make identifier matching (eg cookie syncing) unnecessary for mobile advertising, making them personal data within the meaning of GDPR. On iOS the MAID is called the IDFA and on Android the AAID.

<sup>161</sup> Google Play Developer Policy Center, Monetization and Ads, section on ‘Usage of Android Advertising ID’ available [here](#). The policy states that the terms of use of this ID include: ‘**Respecting users’ selections**. If reset, a new advertising identifier must not be connected to a previous advertising identifier or data derived from a previous advertising identifier without the explicit consent of the user. Also, you must abide by a user’s “Opt out of Interest-based Advertising” or “Opt out of Ads Personalization” setting. If a user has enabled this setting, you may not use the advertising identifier for creating user profiles for advertising purposes or for targeting users with personalized advertising. Allowed activities include contextual advertising, frequency capping, conversion tracking, reporting and security and fraud detection.’

<sup>162</sup> This point was raised in NOYB’s GDPR complaint against Google on the AAID. The complaint states: ‘Studies and official investigations have proved that the AAID is stored, shared and, where needed, linked with old values via countless other identifiers such as IP addresses, IMEI codes and GPS coordinates, social media handles, email addresses or phone number, de facto allowing a persistent tracking of Android users... Far from being a solution, the contractual arrangements with third parties [app developers] are the proof of the technical inability of [Google’s] implementation to comply with the Complainant’s request – eg interrupt the processing of the AAID. The fact that [Google] asks developers not to reconnect the old ID with the new one simply proves that [Google’s] solution does not in fact interrupt processing thus allowing a perpetual tracking of the individual.’ Available [here](#).

<sup>163</sup> Indeed, Google’s EU user consent policy requires publishers and app developers to obtain consent for cookies and mobile identifiers even if they do not serve personalised ads, because non-personalised ads served by Google (Ad Manager, AdSense and AdMob) still use cookies or mobile identifiers to combat fraud and abuse, for frequency capping and for aggregate ad reporting. See Google ‘[Help with the EU user consent policy](#)’.

<sup>164</sup> Disabling pre-installed apps is described [here](#), but uninstalling [requires rooting the phone](#), which ordinary users are not in a position to do, [nor is it safe for them to](#) from a security perspective.

controls that might allow a user to reduce their exposure to trackers. We discuss in more detail the controls and measures in Chrome and Safari, the two browsers with the highest share of use in the UK.<sup>165</sup>

143. Many browsers give users some of the following features that may limit tracking to a limited extent: the ability to clear browsing data (history of sites visited, cookies, and the cache), block the setting of cookies, clear data on a site-by-site basis, browse in private browsing mode, or enable a Do Not Track request to be sent to sites with their HTTP requests. We discuss these in turn next.

144. An overview of browser settings available to users for common browsers is given in Table G.3.

**Table G.3: Browser settings that may limit tracking for Chrome, Safari, Firefox and Edge.**

	<i>Chrome</i>	<i>Safari</i>	<i>Firefox</i>	<i>Edge</i>
Clear history	Yes	Yes	Yes	Yes
Clear cookies	Yes	Yes	Yes	Yes
Clear cache	Yes	Yes	Yes	Yes
Clear data on site-by-site basis	Yes	Yes	Yes	No
Block data storing on a site-by-site basis	Cookies only	No	Yes	No
Block all cookies	Yes	Yes	Yes	Yes
Block third-party cookies	Yes	Yes, by default	Yes	Yes
Block Flash	Yes	Yes, by default	No	Yes
Block JavaScript	Site-by-site basis only	Yes	No	Yes
Device sensors access controls	Yes	Site-by-site basis only	Yes	Yes
See current data stored by site	Yes	Yes	Yes	No
Do Not Track (DNT)	Yes	No	Yes	Yes
Private browsing mode	Yes	Yes	Yes	Yes
Tracking prevention	No	Yes	Yes	Yes

Source: CMA. Data was collected from manual analysis in the latest versions of browsers settings.

145. As shown in Table G.3, most browsers allow some of the data it stores about users' browsing (including data stored by trackers) to be cleared, including history, cookies, cache and sometimes local storage. Most browsers also provide controls for blocking some tracking technologies, such as cookies and JavaScript. Firefox does not have a user-friendly way to block JavaScript.<sup>166</sup> Most browsers will also have settings to block, or always prompt the user before a site tries to access some sensors – including camera, location or

<sup>165</sup> We note that other browsers have also taken tracking prevention measures, which we do not discuss in this appendix. For example, see Firefox's [Enhanced Tracking Prevention](#) which uses a blacklist similar to Edge's [Tracking Prevention](#), or Brave's [measures](#) to tackle fingerprinting. Of particular technical interest, the Tor browser has a [design document](#) that goes into some detail on the technical requirements of cross-origin unlinkability.

<sup>166</sup> It can be done by typing about:config in the address bar in Firefox and setting javascript.enabled to False. However, this is not in the browser settings alone with other settings and not very user-friendly as is not advertised as Firefox's main way to change browser settings.

microphone. Browsers do not generally have controls that allow users to set persistence preferences more generally, eg to set cookies to clear every week automatically, although some have an option to set data to clear when the window is closed.

146. Most browsers offer some form of 'private browsing' mode, although users frequently misunderstand the relatively limited privacy benefits that it provides.<sup>167</sup> When using private browsing, cookies, history and sometimes form data typically stored by the browser will be cleared when the window (the session) is closed. Private browsing does not attempt to mitigate cross-site tracking within a session; it just reduces persistence of data storage, subject to users closing their browser window. Private browsing does not make the user agent (the browser) anonymous on the network either, as the IP address is still sent with every HTTP request as usual. In one study, 56% of surveyed users believed their search queries would not be saved in private browsing mode, even whilst logged into a Google Account, conflating the browser's history with Google's history.<sup>168</sup>
147. Do Not Track (DNT) was a proposed HTTP header field, designed to let users opt-out of tracking in their browser settings. With every HTTP request sent, the DNT request would be sent to websites, and it was hoped that websites would respect this and implement changes for users who had enabled DNT on their browser. DNT was proposed in 2009, and the W3C set up a working group to standardise it in 2015. DNT was implemented by Firefox, Internet Explorer, Safari, and Chrome. However, there was no regulatory requirement for its use and there were no technical measures to enforce users' preferences. Most websites (including Google, Facebook, and Twitter) ignored it.<sup>169</sup> The W3C disbanded the DNT working group in 2019, and Safari dropped support of DNT so that users would not be presented with a misleading and ineffective privacy control.<sup>170</sup>

## Chrome

---

<sup>167</sup> DuckDuckGo, A Study on Private Browsing: Consumer Usage, Knowledge, and Thoughts, January 2017, available [here](#). Wu, Y., Gupta, P., Wei, M., Acar, Y., Fahl, S., & Ur, B. (2018, April). Your secrets are safe: How browsers' explanations impact misconceptions about private browsing mode. In *Proceedings of the 2018 World Wide Web Conference* (pp. 217-226), available [here](#). Habib, H., Colnago, J., Gopalakrishnan, V., Pearman, S., Thomas, J., Acquisti, A., ... & Cranor, L. F. (2018). Away from prying eyes: analyzing usage and understanding of private browsing. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)* (pp. 159-175), available [here](#).

<sup>168</sup> Wu, Y., Gupta, P., Wei, M., Acar, Y., Fahl, S., & Ur, B. (2018, April). Your secrets are safe: How browsers' explanations impact misconceptions about private browsing mode. In *Proceedings of the 2018 World Wide Web Conference* (pp. 217-226), available [here](#).

<sup>169</sup> DuckDuckGo, 'The "Do Not Track" Setting Doesn't Stop You from Being Tracked', June 2020, available [here](#).

<sup>170</sup> Safari's Intelligent Tracking Prevention (ITP) blog [announces](#) dropping support for DNT with the 2.1 release of ITP, which blocks more types of tracking by default for the user.

148. Chrome is the browser with the largest market share (as discussed in Appendix E), and thus the settings, controls and defaults it implements to limit tracking will have a significant impact on tracking practices. There are a few recent developments Chrome have taken in this direction. We discuss the first two in sections on 'Recent and near-future developments in tracking' later in this appendix, and the final development here:
- (a) Chrome's intention to deprecate third-party cookies by 2022;
  - (b) Chrome's Privacy Sandbox developments;
  - (c) Chrome's move prevent third-party access to cookies by default, unless cookies are labelled as being intended for third-party use.
149. Chrome version 80 introduced a new secure-by-default model that assumes all cookies should be protected from third-party access unless web developers specify otherwise. Specifically, only cookies with "SameSite=None; Secure" enabled by the website developer are available for third-party access; all other cookies are limited to first-party access.<sup>171</sup> In the past, developers were not required to label third-party cookies, so Chrome knew only whether a cookie was third-party to a particular page (from the user's context, ie the page they are currently on), not whether it was intended for third-party use generally. By requiring website developers to add the "SameSite=None; Secure" attribute, Chrome can identify third-party cookies without needing to compare the cookie domain to the current page domain. This allows Chrome to offer more precise controls to users who wish to target cookies meant for third-party use. The SameSite attribute is being discussed in the Internet Engineering Task Force (IETF) and adoption plans by different browsers can be viewed online.<sup>172</sup> Safari have not adopted SameSite, perhaps because they have already fully banned third party cookie access as part of Intelligent Tracking Prevention. The lead developer of ITP has criticised SameSite saying, 'developers can simply reconfigure their cookies to opt out of this new policy and we should expect all trackers to do so immediately'.<sup>173</sup>

### *Safari*

150. The Safari browser holds 34% share of use across all device types (desktop, mobile and tablet) in the UK,<sup>174</sup> making it the largest competitor to Chrome. This is mostly from iOS users, where Safari is the default browser. Safari's browser engine is called WebKit, and has in recent years been taking

---

<sup>171</sup> For more information on how this works, see [SameSite cookies explained](#).

<sup>172</sup> CanIUse shows adoption by browser and version, [here](#).

<sup>173</sup> As John Wilander states in a comment [to this tweet of his](#).

<sup>174</sup> According to StatCounter's statistics from May 2020, available [here](#).



measures to combat cross-site tracking. Notably, WebKit launched Intelligent Tracking Prevention (ITP) in 2017.

151. In 2019, Apple complemented ITP by publishing a Tracking Prevention Policy,<sup>175</sup> which sets out its definition of tracking, the types of tracking it will prevent,<sup>176</sup> enforcement, and acknowledged various unintended impacts (including the funding of websites using personalised advertising, and the measurement of the effectiveness of advertising).
  - (a) The policy defines tracking in the same way as we do in this appendix,<sup>177</sup> and aims to prevent all covert and cross-site tracking, without exceptions. It defines a 'privileged third party' as one that can track a user across websites without their knowledge<sup>178</sup> or consent due to special access<sup>179</sup> built into the browser or operating system.
  - (b) WebKit has (with ITP) or intends to implement technical protections to enforce the policy. If WebKit cannot prevent the tracking completely without causing undue user harm it will limit it; for example, by reducing persistence or available bits of entropy, or relying on alternative technologies.<sup>180</sup> If this is not possible, WebKit will ask for the users consent as in Figure G.9.<sup>181</sup>
  - (c) Apple states that circumvention is treated with the same seriousness as exploitation of security vulnerabilities. Apple states that when faced with a trade-off arising from unintended impacts of its policy, it will typically prioritise user benefits over preserving current website practices, stating its belief that that is the role of the web browser as the user agent.
152. Intelligent Tracking Prevention (ITP) is comprised of two main stages:
  - (a) A machine learning classifier to distinguish which domains are trackers; and

---

<sup>175</sup> Available [here](#) on their blog.

<sup>176</sup> They categories types of tracking including cross-site tracking, stateful tracking, covert stateful tracking, navigational tracking, fingerprinting and covert tracking. We have not used many of these terms in this appendix but the reader is referred to [their blog post](#) which explains them.

<sup>177</sup> WebKit [define tracking](#) as the collection of data regarding an individual's identity or activity across one or more websites. Even if this data is not believed to be personally identifiable, it is still tracking. This definition is similar to our definition from Disconnect. Apple make a similar point that we made in the section on 'Linking identifiers', tracking does not require an individual identifier to be personal data for it to be used in combination with others to identify an individual.

<sup>178</sup> Interactions are considered third-party even if the user is transiently informed in context. For example, a redirect that is triggered by the user hovering over or muting some content.

<sup>179</sup> Pre-installed apps (discussed earlier in the section on 'Pre-installed apps') are an example of privileged software.

<sup>180</sup> Such as the [Storage Access API](#) or [Private Click Measurement](#).

<sup>181</sup> WebKit [considers](#) some actions, like using single sign-on (SSO), to be implied consent if these logins are noticeable and active on the user's part.

(b) A set of implementations to curtail/limit the ability of trackers once classified as such.

153. The data input to ITP's classifier includes statistics on user interactions (events) and resource loads by domain. The classifier learns whether a domain is tracking the user largely based on features found to be predictive.<sup>182</sup> Patterns are learnt over time, and by ITP 2.0 the classifier had also learnt to identify tracker collusion<sup>183</sup> and first-party bounce trackers.<sup>184</sup> This machine learning approach can be contrasted with more simplistic methods, such as those that rely on blacklists of tracking domains that are manually maintained. In principle, ITP's classifier can continuously learn a model of what a tracker looks like in general, and detect and add new domains to its list automatically based on this model.
154. Once a domain has been classified as having the ability to track the user cross-site, various restrictions on cookies and their persistence are enforced by Safari. Apple has been increasing the restrictions and improving ITP's features since 2017, in response to adtech workarounds, and are doing more than just cookie blocking. We do not give a full history of these developments, but summarise the current situation:<sup>185</sup>
- (a) All third-party cookie setting is blocked automatically;
  - (b) First-party cookies cannot be read in a third-party context, impacting attribution. As an alternative, WebKit encourages use of the Storage Access API,<sup>186</sup> which prompts the user for consent, putting Safari in a stewardship role (see Figure G.9).
  - (c) All client-side first-party cookies<sup>187</sup> expire after seven days. If classified by ITP as a tracking domain, client-side first-party cookies expire after 24 hours.

---

<sup>182</sup> For example, in [ITP 1.0](#) the features found to be most predictive of a tracker included subresource under number of unique domains, sub frame under number of unique domains, and number of unique domains redirected to.

<sup>183</sup> In tracker collusion, one tracker passes a message 'this is user ABC' to another tracker ad infinitum, discussed [here](#) in the blog on ITP 2.0.

<sup>184</sup> These are when a user is quickly redirected to a tracking domain in the background before being passed to their destination, explained [here](#).

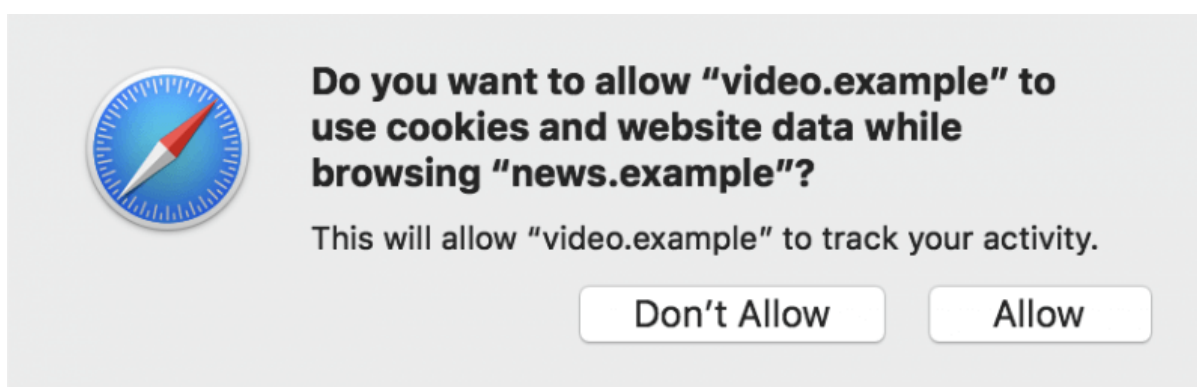
<sup>185</sup> Clearcode give a summary of ITP available [here](#).

<sup>186</sup> This API was made by WebKit to allow a way for embedded content (such as iFrames) to still function when users are both authenticated and consenting. The Storage Access API is discussed in depth [here](#).

<sup>187</sup> Client-side cookies are those created by JS on document.cookie which override those set via the server using the HTTP Set-Cookie attribute, as discussed earlier in this appendix in the section on 'Cookies'. Trackers often set cookies client-side because they themselves are client-side (they are JavaScript tags/snippets and execute in the browser). This ITP restriction does not apply to cookies set by the web server, as this indicates a more deliberate decision from the site owner than including a JavaScript tag.

- (d) Crackdown on domains using HTTP redirects, link decoration and the Referrer header in the background for tracking. If detected, associated cookies are purged, tracking domains blacklisted and URLs are truncated. For example, "https://sub.social.com/some/path/?clickID=012345678" was truncated to "https://sub.social.com" by ITP 2.0 and to "https://social.com" by ITP 2.3.
- (e) Limits HSTS<sup>188</sup> state to address the abuse of HSTS cache for storing a unique identifier.<sup>189</sup>
- (f) A seven-day expiry on data in a selection of non-cookie browser stores, including localStorage.

**Figure G.9: An example of ITP prompting the user about third-party trackers.**



Source: Webkit blog announcing ITP 2.0 available [here](#).

155. ClearCode break down the impact of ITP on companies, publishers, analytics and walled-garden ecosystems in their blog.<sup>190</sup> It notes that adtech vendors such as SSPs, DSPs and DMPs will not be able to set third-party cookies, meaning they will need to find alternatives for targeting and attribution.<sup>191</sup> Also, if they get classified as a tracker by ITP, the persistence of some alternatives is curtailed.<sup>192</sup> For publishers, the impact is felt through a drop in average prices (CPM) for their display advertising inventory, due to advertisers being unable to identify the user on the website and therefore bidding less for an impression.<sup>193</sup> Analytics companies are affected too, as a 7-day persistence is enforced for first-party cookies, which limits metrics and reporting for Safari users if their previous visit was more than seven days ago.

<sup>188</sup> HSTS (HTTP Strict Transport Security) is a web security policy mechanism that protects sites from multiple attacks by enforcing the client (browser) request to only be dealt with if coming via a secure connection (using SSL, ie. HTTPS only). More information is available from OWASP [here](#). See also [this blog](#) that discusses an abuse of HSTS using wildcard SSL certificates to store, edit and retrieve data in users' browsers.

<sup>189</sup> This measure is described in more depth in WebKit's blog post [here](#).

<sup>190</sup> In ClearCode's blog post, available [here](#).

<sup>191</sup> Apple has proposed [Privacy Preserving Ad Click Attribution](#) to solve this, and this is discussed in the section on privacy-enhancing technologies below under 'WebKit Private Click Measurement'.

<sup>192</sup> To 24 hours for first-party cookie setting, and to 7 days for data in localStorage.

<sup>193</sup> See Appendix F for our analysis of an RCT that Google performed to measure the impact of removing the information associated with third-party cookies on publishers' revenues.

Until ITP 2.0, the impact on walled gardens was minimal, however this changed when the enforcement to get user consent using the Storage Access API prompt began (Figure G.9).

156. When the first version of ITP was released, six major advertising trade associations, including the IAB and NAI, wrote a joint open letter ‘from the Digital Advertising Community’ in opposition to ITP, criticising it as ‘unilateral and heavy-handed’.<sup>194</sup> This illustrates the significant impact that browsers can have on the digital advertising market.
157. In sum, Apple has taken significant steps with ITP and caused sizeable impact in the adtech industry. It is still issuing releases to update ITP. The arms race between Apple and adtech firms trying to work around ITP continues.<sup>195</sup>

### **Internet Governance**

158. Standards are agreed-upon technical specifications that underpin the infrastructure of the internet.<sup>196</sup> Standardisation is a powerful tool, given that the web works best when there is compatibility and interoperability between protocols and applications. Governance and standard setting on the web occurs mainly via two routes: (i) platforms such as browsers or mobile OSs may implement new features or approaches, and encourage other platforms to follow suit, or alternatively (ii) Standards Setting Organisations (SSOs) may develop a common standard and try to get platforms to adopt it. SSOs tend to be comprised of volunteers, and usually operate under some form of consensus decision-making.
159. Some SSOs relevant to user tracking and other issues in this market study include:
  - (a) the Internet Engineering Task Force (IETF);
  - (b) the World Wide Web Consortium (W3C); and
  - (c) the Web Hypertext Application Technology Working Group (WHATWG).
160. In this section, we do not attempt to cover the full history and governance structures of SSOs, nor to provide a complete list of relevant SSOs. Instead, we select some of the most prominent ones (listed above) and briefly examine the key technologies relevant to tracking that these SSOs work on.

---

<sup>194</sup> Six major advertising associations wrote to Apple as reported on by Adage available [here](#).

<sup>195</sup> One of the better technical blog posts on this is by Simo Ahava, available [here](#).

<sup>196</sup> As the Internet Society define in their policy brief on open internet standards, available [here](#).

161. The IETF is an open standards organisation which develops and promotes voluntary internet standards, most notably those comprised in the Internet protocol suite (TCP/IP). The IETF describes itself as a large open international community, with various working groups that fall into Areas,<sup>197</sup> with much of the work being handled via mailing lists.<sup>198</sup> It was established in 1986 with support from the US federal government, but since 1993 operates independently under the Internet Society, an international membership-based not-for-profit.
162. Whilst the IETF's focus is on the backbone of the internet (TCP), the Internet Society is concerned with several issues that may be relevant to consumer and competition authorities. For example, of relevance to this appendix, they publish short Policy Briefs on topics such as privacy and identity on the internet.<sup>199</sup> One Policy Brief describes 'Internet Invariants', properties they believe should not change even as the internet does.<sup>200</sup> Those particularly relevant to consumer and competition are 'Accessibility' (that anyone can access the internet to consume and contribute content) and 'No permanent favourites' (the idea that good ideas will be overtaken by better ones and removing competition is standing in the way of the internet's natural evolution).<sup>201</sup>
163. The IETF most notably looks after the Transport Communication Protocol (TCP) and UDP (User Datagram Protocol) which underly HTTP, the communication protocol of the web. It also looks after HTTP,<sup>202</sup> which as discussed earlier has several features that can be used in tracking, including HTTP headers and link decoration. The IETF has also set up working groups on the Internet of Things (IoT),<sup>203</sup> which is likely to be increasingly of interest to consumer protection regulators who might assess the user data accessed by device sensors, as discussed earlier.
164. Also of relevance to tracking as discussed earlier, the Request for Comments (RFC)<sup>204</sup> that defines the Web Origin Concept was written by the IETF.<sup>205</sup> This defines the Same-origin Policy (SOP), discussed earlier, which restricts how a script or document loaded from one origin can interact with a resource from

---

<sup>197</sup> The IETF's areas are listed [here](#).

<sup>198</sup> As the IETF's about page mentions [here](#).

<sup>199</sup> A full list of Policy Briefs published by the Internet Society is available on their website [here](#).

<sup>200</sup> The full list of these are available in the Internet Society's Policy Brief on Internet Invariants, available [here](#).

<sup>201</sup> As described by the Internet Society [here](#).

<sup>202</sup> HTTP was originally initiated by Tim Berners-Lee at CERN in 1989, and early development of HTTP was coordinated by together by the IETF and W3C, with work later moving to the IETF.

<sup>203</sup> See the full list of IoT related topics being worked on at the IETF [here](#).

<sup>204</sup> RFCs are documents created by the IETF and other Internet Society organisations which are authored by engineers and computer scientists in the form of a memorandum describing methods, research or proposals. The IETF adopts some RFCs as Internet Standards, although many are considered experimental in nature.

<sup>205</sup> RFC6454 on the Same Origin Policy is available [here](#).

another origin.<sup>206</sup> The SOP essentially restricts cross-site tracking (as well as various attacks such as cross-site request forgery), and is thus of great importance in protecting users online.

165. In addition to the IETF, a notable SSO is the W3C. The W3C is an international standards organisation for the web. It was founded in 1994 by Tim Berners-Lee, and is comprised of member organisations who have full-time staff working in various working groups. The W3C has relevant working groups including one on privacy, one on tracking prevention (which closed recently),<sup>207</sup> and one on improving web advertising.<sup>208</sup>
166. Notably, the W3C historically looked after the Document Object Model (DOM), the programmatic interface to a web page (HTML) which allows programs to dynamically access and update the content of webpages. This is of great relevance to how tracking works, given that most web trackers take the form of JavaScript tags or pixels that may collect user data from the webpage via the DOM.
167. Furthermore, the W3C has also defined a number of significant JavaScript Web APIs.<sup>209</sup> Of particular note is XMLHttpRequest API, which allows for HTTP requests to be made dynamically with JavaScript and is one way that trackers may send a user's data back to their domains. Another Web API maintained by the W3C is the Geolocation API which makes the user's current location available to browser-based applications. Notably, the IETF wrote to the W3C about privacy concerns of the Geolocation API,<sup>210</sup> in an example of how SSOs advise each other.
168. In 2004 a group of individuals from leading web browser vendors came together into the WHATWG, which was formed in response to the slow development of W3C web standards including the W3C's decision to abandon HTML in favour of XML.<sup>211</sup> WHATWG introduced HTML5, known now as the HTML Living Standard which comprises a number of technologies that are used in tracking such as web storage (including localStorage)<sup>212</sup> and web

---

<sup>206</sup> Two URLs have the same origin if the protocol, port and host are all the same for both.

<sup>207</sup> The [TPWG](#) (Tracking Prevention Working Group) was overseeing the failed Do Not Track initiative, discussed earlier in the section on Safari. The TPWG is currently closed as of January 2019.

<sup>208</sup> The W3C group for improving web advertising is [here](#) and an article by Digiday on it is [here](#).

<sup>209</sup> The full list of Web APIs that the W3C are defining is [here](#).

<sup>210</sup> Several privacy concerns were raised in the letter from the IETF to the W3C on the Geolocation API, available [here](#).

<sup>211</sup> As the history of HTML written [here](#) specifies.

<sup>212</sup> Web storage is also known as DOM storage and provides web apps with methods for storing data client-side (on the browser). Web storage supports persistent data storage and includes localStorage and sessionStorage. These behave similar to persistent cookies and session cookies respectively.



workers.<sup>213</sup> WHATWG also took over maintaining standards from W3C for the XMLHttpRequest API from W3C in 2012,<sup>214</sup> and the DOM in 2004.<sup>215</sup>

169. Such changes in oversight on standardisation development suggest that this ecosystem is dynamic and undergoes continuous development. Ultimately, any standard will need to be adopted by browsers and other platforms in order to be implemented. Historically, many successful collaborations in setting standards between browsers and SSOs can be cited. For example, JavaScript was developed initially by Netscape (the predecessor of Firefox) in conjunction with Ecma International.<sup>216</sup> Additionally, the SOP was initially developed by Netscape, after which the IETF took it forward for standardisation.
170. In the mobile ecosystem, the role of SSOs has been less prominent.<sup>217</sup> The majority of the defining moves have been implemented by the most prominent mobile OSs, iOS (Apple) and Android (Google). As discussed in previous sections, mobile OSs play a central role in managing access to user data, including the permissions model for access to device-generated data, developer agreements for apps that publish to the play/app store, and (for Android, which is open source) agreements with OEMs who may facilitate the tracking of users by pre-installed apps.
171. In sum, SSOs and platforms tend to work together to establish norms and standards on the internet. Some of the technologies they maintain are very relevant for users' overall welfare, such as those that might enable or restrict the ability of trackers. We note that SSOs are more prevalent in the internet/web space than in the mobile ecosystem (although there is some overlap).

## Tracking in digital advertising

172. The display advertising intermediation ecosystem is complex. The core participants and value chain are discussed more fully in Appendix M.
173. This section focuses on the applications of the technologies and ideas explained in the previous sections within adtech. It covers:

---

<sup>213</sup> Web workers are a way for web content to run scripts in background threads,

<sup>214</sup> As specified by WHATWG [here](#).

<sup>215</sup> The DOM specification is given [here](#) by the WHATWG.

<sup>216</sup> [Ecma International](#) was first established in 1961 under its old name ECMA (European Computer Manufacturers Association).

<sup>217</sup> There is one exception to this, which is the Open Mobile Alliance who worked on the Wireless Application Protocol which was used for mobile markup language before most mobile internet browsers supported HTML, which they do now.



- (a) the role of consent in digital advertising, the technologies involved in the use of data and tracking in the adtech ecosystem for programmatic display advertising and real-time bidding (RTB), and the data protection concerns with RTB;
- (b) cookie matching, a method by which different parties in the adtech ecosystem achieve common identification of users, which is often necessary for intermediaries to exploit and share data about users;
- (c) the role of data brokers and data management platforms (DMPs) in the adtech ecosystem in facilitating cross-site tracking and linking data about a person across multiple contexts and properties owned by different data controllers;<sup>218</sup> and
- (d) the use of tracking technologies, and the use and transfer of data about users, by Google and Facebook’s advertising and analytics services.

### ***Programmatic display advertising and real-time bidding (RTB)***

174. Most digital advertising is sold ‘programmatically’, using automated systems and processes to buy and sell inventory in real time. When users visit a webpage or app with an ad space, the publisher may send a request for bids from advertisers to fill that space, using protocols like the IAB’s OpenRTB<sup>219</sup> and Google’s Authorized Buyers RTB.<sup>220</sup> RTB exchanges collectively send out many billions of bid requests every day.

### ***Lawful basis for processing personal data in RTB***

175. The ICO has stated that ‘the nature of the processing within RTB... means that legitimate interests cannot be used for the main bid request processing... the only lawful basis for “business as usual” RTB processing of personal data is consent (ie processing relating to the placing and reading of the cookie and the onward transfer of the bid request)’.<sup>221</sup>

176. We note that ICO has observed that ‘at present, some parts of the adtech industry are unaware of this advice’.<sup>222</sup> One DSP told us that, in its experience, ‘explicit consumer consent is present for 50-60% of EU ad impressions... using the IAB Transparency and Consent Framework. The remaining 40-50% of traffic are publishers which do not transmit consent

<sup>218</sup> The role of DMPs is also discussed with in Appendix M.

<sup>219</sup> OpenRTB (Real-Time Bidding) 3.0, available [here](#).

<sup>220</sup> Google Authorized Buyers Real-Time Bidding Proto, available [here](#).

<sup>221</sup> ICO, Update report into adtech and real time bidding, 20 June 2019, section 3.3, available [here](#).

<sup>222</sup> ICO, Update report into adtech and real time bidding, 20 June 2019, section 3.3, available [here](#). [ibid.]

signals, quite often because they regard user consent as not necessary and operate under legitimate interest.’

### *Transparency and Consent Framework (TCF)*

177. Compliant websites and apps should obtain valid user consent before setting cookies and triggering tags or pixels to send bid requests.<sup>223</sup> In addition, information about the user’s consent or objections should be communicated to adtech participants within the set of providers that the publisher has chosen to work with.
178. One prominent method to do so is IAB Tech Lab and IAB Europe’s Transparency and Consent Framework (TCF), which relies on consent management platforms (CMPs). CMPs are entities that help publishers to centralise and manage (i) the presentation of information to users about the purposes, features and legal bases for processing personal data (‘transparency’), and (ii) acquire consent and manage objections from users (‘consent’).
179. CMPs often present the user interface (UI) by which transparency and consent are achieved when users first visit a webpage or app, or when any previously set consent preferences stored in cookies are deemed to have expired. Figure G.10 illustrates with a stylised example. CMPs communicate this information on behalf of the publisher to the rest of the ecosystem using standardised strings of numbers included in bid requests (called ‘Transparency and Consent Strings’, or ‘TC Strings’) and a standard API to create and process TC Strings.<sup>224</sup>

**Figure G.10: a stylised example of a CMP UI**



Source: Econsultancy, [What is a consent management platform, and are they needed?](#)

<sup>223</sup> Consent must always be obtained before the controller starts processing personal data for which consent is needed. (Article 29 Working Party, Guidelines on Consent under Regulation 2016/679, pp.17-18, available [here](#).)

<sup>224</sup> IAB Tech Lab, Consent Management Platform API, available [here](#).

180. IAB Europe maintains a list of registered CMPs.<sup>225</sup> It also maintains a Global Vendor List (GVL), which is a list of all registered and approved adtech third parties ('Vendors') participating in the TCF. Furthermore, it sets out policies which govern how TCF should be used as well as a list of standard purposes and features of data processing for adtech.<sup>226</sup> At the time of writing, the TCF vendor list comprises over 520 organisations,<sup>227</sup> and the list of registered TCF v2.0 vendors has over 390 entities.<sup>228</sup> Publishers can control which adtech providers to work with and configure their CMP to only request consent for and send information to those providers.<sup>229</sup>
181. Whilst publishers can choose not to use TCF, TCF appears to have established or is close to establishing critical mass to become an industry-wide standard, particularly as Google has stated that Google Ad Manager (its publisher ad server and SSP) will be integrating with TCF v2.0.<sup>230</sup>
182. However, we note some concerns with TCF that have been expressed:
- (a) Many websites have TCF user interfaces that may not be compliant with GDPR, designed to nudge users to take a particular course of action. The ICO has stated such practices are non-compliant.<sup>231</sup> Nouwens et al. (2020) scraped the interface designs of the five most popular CMPs on the top 10,000 websites in the UK.<sup>232</sup> The authors looked for whether (i) consent is explicit (ie a clear, positive, affirmative act such as clicking a button, rather than eg continuing to navigate a website); (ii) accepting all is as easy as rejecting all (in terms of number of clicks required); and (iii) no pre-ticked boxes (ie no non-necessary purposes or vendors are pre-selected to be on). These three conditions, which are more readily measurable using the authors' methodology, are necessary but not sufficient conditions for compliance with GDPR. The authors found that only 11.8% of the scraped websites met all three requirements.

---

<sup>225</sup> IAB Europe, CMP List, available [here](#).

<sup>226</sup> The purposes and features of processing, along with standardised legal and user-friendly text, and guidance for vendors, are listed the appendices of the IAB Europe Transparency & Consent Framework Policies, available [here](#).

<sup>227</sup> IAB Europe, Vendor List, available [here](#).

<sup>228</sup> IAB Europe, Vendor List TCF v2.0, available [here](#).

<sup>229</sup> Google offers publishers using Google Ad Manager similar controls about which adtech providers on Google's own whitelist can work with the publisher's traffic in the EEA and UK. It states, in its Ad Manager and Ad Exchange program policies (available [here](#)), that if publishers don't engage with these controls, it will apply a default of nearly 200 commonly used adtech providers.

<sup>230</sup> Google Ad Manager, Ad Manager and Ad Exchanged program policies IAB Transparency and Consent Framework v2.0, available [here](#).

<sup>231</sup> See ICO, Guidance on the use of cookies and similar technologies, available [here](#) – in particular, the section on 'Can we pre-enable any non-essential cookies?' which sets out the ICO's view on an example where a consent mechanism that emphasises 'allow' over 'do not allow'.

<sup>232</sup> Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. *arXiv preprint arXiv:2001.02479*. Available [here](#). In the paper, 680 of the 10,000 websites contained a CMP that could be scraped using a purpose-built tool.

- (b) Similarly, the Irish DPC published a report in April 2020 of a sweep of 40 popular websites to examine the use of cookies and similar technologies, and in particular how data controllers obtain the consent of users for these tracking technologies. The DPC stated that ‘almost all of the sites continue to have compliance issues, ranging from minor to serious’.<sup>233</sup>
- (c) A proportion of bid requests involve processing of special category data (either directly or by inference), such as bid requests for webpages or apps related to politics, religion, ethnic groups, health, etc. In the ICO’s view, there is a general prohibition on processing special category data as per article 9(1) of the GDPR, unless one of the specific article 9 conditions for processing apply, of which ‘explicit consent’ is the only applicable condition in the context of RTB, a higher standard than ‘regular’ consent.<sup>234</sup> The ICO has stated that ‘the current consent requests provided under both the TCF and [Google’s Authorized Buyer] frameworks are non-compliant’ and do not meet the appropriate standard of explicit consent required for processing special category data.<sup>235</sup>

183. Furthermore, there are doubts about the extent to which websites correctly apply TCF and obtain valid consent:

- (a) Trevisan et al. (2019) found that 49% of a large set of 35,862 websites set cookies before any user consent is given.<sup>236</sup>
- (b) Analysing the consent strings stored in cookies, Matte et al (2019) found that 141 of 1,426 automatically crawled European websites with TCF CMPs (9.9%) set cookies recording positive consent even when the user has not made any choice. Matte et al (2019) further found, using a semi-automatic review of 560 websites with TCF CMPs, that 27 (4.8%) set a cookie for positive consent even if the user refuses consent.<sup>237</sup>

184. Finally, we note that TCF, like most efforts to communicate data protection information to users, is based on a ‘notice and consent’ model, in that users are typically presented with privacy policies and are asked to accept or reject. As detailed in Chapter 4 and Appendix K, most privacy policy notices take a long time to read and often require advanced reading comprehension abilities.

---

<sup>233</sup> Irish Data Protection Commission, Report by the DPC on the use of cookies and other tracking technologies, available [here](#).

<sup>234</sup> Article 29 Working Party, Guidelines on Consent under Regulation 2016/679, section 4, available [here](#).

<sup>235</sup> ICO, Update report into adtech and real time bidding, 20 June 2019, section 3.2, available [here](#).

<sup>236</sup> Trevisan, M., Traverso, S., Bassi, E., & Mellia, M. (2019). 4 years of EU cookie law: Results and lessons learned. *Proceedings on Privacy Enhancing Technologies*, 2019(2), 126-145. Available [here](#).

<sup>237</sup> Matte, C., Bielova, N., & Santos, C. (2019). Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe’s Transparency and Consent Framework. *arXiv preprint arXiv:1911.09964*. Available [here](#).

Users rarely read privacy policies before using a website or app.<sup>238</sup> Many users consent, almost by reflex, without viewing privacy policies, in order to quickly remove the obstacles to their primary goal of accessing the service.<sup>239</sup> TCF, like other ‘notice and consent’ models, places a burden on people to make choices that they are not well placed to make.<sup>240</sup>

### *Programmatic display advertising and real-time bidding (RTB) as a source of data leakage*

185. Most bid requests contain pieces of personal data, which can be used to identify a person, directly or indirectly, either by themselves or in combination with other information that adtech data controllers may have access to.<sup>241</sup>

Some examples are:

- (a) a unique identifier for the bid request (ie a query ID or an auction ID) generated by the SSP;<sup>242</sup>
- (b) identifiers, including cookie IDs, MAIDs, and other matched IDs from third parties which can be used to link together other data about the user;
- (c) a User-Agent string which, as discussed above, contains highly specific information about the user’s browser and OS, and which can be used to fingerprint devices;
- (d) the user’s IP address, location data (such as GPS coordinates), time zone; and
- (e) device information (such as make, model, screen size, detected language of the user’s system, etc.) which can be used to fingerprint devices.

186. Bid requests also contain information on the webpage URL or app that the user is currently viewing.<sup>243</sup> On the one hand, information on the context of

---

<sup>238</sup> See, for instance, Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1), 128-147, available [here](#).

<sup>239</sup> See, for instance, our findings on the low amount of time that users spend on Google and Facebook’s privacy policies (detailed in Chapter 4).

<sup>240</sup> Cate, F. H. (2010). The limits of notice and choice. *IEEE Security & Privacy*, 8(2), 59-62. Slides available [here](#). See also Wired, ‘We need to fix GDPR’s biggest failure: broken cookie notices’, available [here](#).

<sup>241</sup> In contrast to the US concept of ‘personally identifiable information (PII)’, the GDPR includes ‘online identifiers’ within the definition of personal data. See ICO, ‘What is personal data?’, available [here](#).

<sup>242</sup> Note that there is currently no industry-wide common impression ID, which can be used to identify the same impression being sold using multiple SSPs (holding auctions of auctions). Such an ID would enhance transparency for advertisers and help an advertiser and/or DSP better manage the possibility of competing with itself (self-competition) for the same impression through different supply paths. However, it would also facilitate pooling of user data contained in bid requests from multiple SSPs.

<sup>243</sup> Google Ad Manager used to provide its contextual categories (or publisher verticals, as described in [Google’s RTB documentation](#)) for the page that the user is on in its bid requests, in addition to the URL, which bidders could theoretically use to work out the content of the webpage when formulating its bid. These were deprecated in February 2020. However, this likely had a minimal effect on privacy or data protection, since bidders still had

the ad impression allows advertisers to assess its quality and to ensure that their ads do not appear on inappropriate websites or apps (brand safety) when deciding whether to make a bid. On the other hand, contextual information within bid requests is also personal data, as it is about the browsing behaviour of a person that can be identified using the other information contained in the bid request. This contextual information may even constitute special category data where the context is related to or could support inferences about health, sexuality, politics, religion, or ethnicity.

187. These bid requests are sent to potentially hundreds of adtech intermediaries and advertisers, particularly for open auctions, where any advertiser can bid for the impression. Indeed, advertisers and adtech intermediaries do not even need to bid on any impressions in order to create user profiles of browsing histories, simply by receiving bid requests and recording identifying information and URLs (ie 'listening to the bidstream').<sup>244</sup> The ICO has expressed concern that it is not possible for consumers to provide valid consent to this large-scale data processing, with their personal data (potentially including special category data) shared with an unknowable (from the perspective of the consumer) and large number of parties, with unknowable controls and security measures.<sup>245</sup>
188. Currently, publishers may have an incentive to transmit user identifiers in bid requests because, as discussed in Appendices F and M, doing so tends to increase the number and value of the bids they receive in response. However, as discussed in Appendix M, doing so could also lead to 'audience arbitrage' or 'commoditisation' of publishers' audiences, as the data within bid requests is often sufficient to enable advertisers and adtech providers to reidentify users on other websites with cheaper inventory.
189. The data leakage in RTB could also occur in the other direction, from advertisers to publishers. For instance, a publisher may observe that bid requests for a particular user on its website tend to elicit high value bid responses, for instance, for ads relating to niche dating services, addiction treatment, debt servicing, or retargeting ads for baby products. The publisher

---

access to the URL and potentially look up the content of the URL themselves or using an alternative contextual data provider. (See, for instance, AdExchanger, 'Industry shrugs as Google announces plans to restrict contextual data', available [here](#).)

<sup>244</sup> See, for example, Digiday, 'We get audience data at virtually no cost': Confessions of a programmatic ad buyer, available [here](#).

<sup>245</sup> On the risk of data leakage from this process, the ICO stated that 'there are no guarantees or technical controls about the processing of personal data by other parties, eg retention, security etc. In essence, once data is out of the hands of one party, essentially that party has no way to guarantee that the data will remain subject to appropriate protection and controls.' (ICO, Update report into adtech and real time bidding, 20 June 2019, section 3.5, available [here](#).)

may be able to draw inferences about and build up its own profile of that user's demographics and interests.<sup>246</sup>

### ***Cookie matching***

190. Cookie matching, also known as cookie syncing, is a common method by which different parties in the adtech ecosystem achieve common identification of users (or more specifically, their browsers). This is often necessary for intermediaries (like DSPs, and measurement and attribution providers) to be able to use any information they have, and share information about users or browsers with other intermediaries.
191. Cookie matching is necessary because browsers prevent domains from reading cookies set by another domain (the Same Origin Policy). This means that, without cookie matching, different independent adtech intermediaries can only set and read their own cookie identifiers for users' browsers.<sup>247</sup>
192. It is important to note that the cookies in the process described in this section are typically third-party (or cross-domain) cookies. (The distinction between third- and first-party cookies is explained in the section above on 'Cookies'.) Many major browsers are taking more aggressive steps to restrict third-party cookies – for instance, they are blocked by default on Safari and Firefox, and Chrome has announced that they will phase out support for third-party cookies soon. As set out below and elsewhere in this appendix and the report, the third-party cookie is currently a fundamental building block of open display advertising, due to its role in allowing advertisers and publishers to achieve common identification of users. The implications of the removal of third-party cookies on display advertising are discussed in more detail in Chapter 5, Appendix M, and the section on 'Control over web standards and relative dependence of third-party cookies' below.

### ***How cookie matching works***

193. The exact details of cookie matching in an RTB context depends on exactly which combination of independent adtech entities need to establish a sync. We illustrate the general ideas with a single stylized example of a new, unknown user visiting a webpage which is using an integrated publisher ad

---

<sup>246</sup> As explained in Appendix M, Google Ad Manager recently took measures to prevent publishers from linking data about winning and losing bids in their Bid Data Transfer files with impression-level data in Impression Data Transfer file and also other report files, ostensibly to prevent this possibility of publishers being able to link data to user IDs not just from winning ads but also all the losing bids. However, these measures have reduced transparency for publishers and inhibited their ability to assess relative performance of SSPs.

<sup>247</sup> Cookies, the Same Origin Policy, and the general idea of linking identifiers is explained in their respective sections above. URL redirects are explained in the section on Hypertext Transfer Protocol (HTTP).



server and exchange/SSP (such as Google Ad Manager and Authorized Buyers).<sup>248</sup>

194. When the user visits a website (and provides the appropriate consent), the publisher's integrated ad server and SSP sets a cookie ID for the user's browser. The integrated SSP sends bid requests to potential buyers. Suppose that one buyer/DSP, using only the information contained in the bid request, makes a successful bid. The DSP would send an ad containing tags that enables the DSP to set its own cookie ID for that user's browser, and a match tag (provided to the DSP by the SSP) that makes a request to the SSP's cookie matching service. This request would either i) contain the DSP's cookie ID, which the SSP could link to its own cookie ID (if the SSP is storing the match table), or ii) request the SSP to send its cookie ID to the DSP, which the DSP could link to its own cookie ID (if the DSP is storing the match table),<sup>249</sup> or iii) both (if the match is symmetrical and both parties are storing a match table).<sup>250</sup>
195. In this way, both the winning DSP and the SSP would know how the other refers to that user. In future, if that user visits a webpage which leads to the same SSP sending bid requests to the same DSP, the SSP could include its cookie ID and/or the DSP's cookie ID in the bid request, allowing the DSP to look up that user and make use of any information it has on that user when formulating its bid response (including choosing ads that are personalised to that user).

#### *Piggybacking and sharing match tables*

196. Cookie matching is not limited to the winning DSP and SSP. The SSP may redirect the user's browser to multiple other DSPs and DMPs that participate on its exchange (and which the publisher has approved<sup>251</sup> and the user has given 'consent' for). Through this redirection, it gives those other intermediaries an opportunity to set their own cookie IDs and sync it with the SSP's cookie ID. Similarly, the winning DSP may redirect the user's browser to other SSPs that the user has 'consented' to.

---

<sup>248</sup> An additional match may be required for publisher ad servers interacting with third-party SSPs (eg a publisher using Google Ad Manager making use of Open Bidding).

<sup>249</sup> Alternatively, the DSP may request the exchange to send the exchange's cookie ID to another third-party, such as a DMP that the DSP is using to host match tables.

<sup>250</sup> In the case of Google Ad Manager and Authorized Buyers, Google sends an encrypted version of its cookie ID that is specific to each buyer. See Google's Authorized Buyers Real-Time Bidding documentation on Cookie Matching, available [here](#).

<sup>251</sup> For example, Google states that its piggyback requests (which it calls 'pixel matching') '[do] not operate on the properties of publishers who opt out of the additional match.' (See section on 'Pixel Matching' in Google's documentation on Authorized Buyers Real-Time Bidding Cookie Matching, available [here](#).)

197. Google, for instance, told us that it may initiate cookie matching to third-party buyers, even when that buyer did not serve an ad, in order to increase the number of links and to assist buyers that may have relatively few opportunities to serve a match tag. Google states that it places limits and restrictions to protect user privacy, such as only selecting one additional buyer to send a piggyback call (what it refers to as ‘pixel match’),<sup>252</sup> prohibiting multiple buyers from joining data they receive from its cookie matching service,<sup>253</sup> and prohibiting the use of its cookie matching service for the purpose of data harvesting.<sup>254</sup>
198. Indeed, a prominent DMP told us that, in contrast to the first-party data uploaded by its customers and which are only available to that customer, the DMP (and in some cases other third-party adtech and data providers) shares match tables across all its customers, so that ‘each customer can benefit from the match tags being fired on other customers’ websites’.

### *Cookie matching is a real-time process*

199. Cookie matching relies on a HTTP redirect, so it is done in real-time by the browser when the user is on the publisher’s website. Typically, the cookie matching process occurs whilst the user is on the website, but after the main content of the webpage has loaded and not before,<sup>255</sup> as otherwise the number of matches could lead to high latency and long load times. Nevertheless, ad loading is slow if there are many syncs,<sup>256</sup> and ads and pages that take too long to load may lose the opportunity to be viewed by users (let alone convert) if they abandon the webpage before ads finish loading.

---

<sup>252</sup> This is described in the section on ‘Pixel Matching’ in Google’s documentation on Authorized Buyers Real-Time Bidding Cookie Matching, available [here](#). ‘In cookie matching, the buyer that wins the auction for an impression can associate a cookie with a Google User ID. In another component of Google’s cookie matching code, called pixel matching, Google algorithmically selects an additional buyer whose cookie can be matched with the Google User ID. Google then places a match tag onto the impression, and includes the chosen buyer’s URL in the match tag.’

‘Pixel matching does not operate on the properties of publishers who opt out of the additional match.’

<sup>253</sup> ‘The Cookie Matching Service respects user privacy by adhering to the following principles: [...] Google prohibits multiple buyers from joining data they receive from the Cookie Matching Service. [...] The purpose of the match table is to allow buyers to use the information they own about the user in transacting with Google. The use of the Cookie Matching Service for the purpose of data harvesting is strictly prohibited by the Authorized Buyers contract and policies.’ (Google’s documentation on Authorized Buyers Real-Time Bidding Cookie Matching, section on ‘Respects user privacy’, available [here](#).)

<sup>254</sup> *ibid*.

<sup>255</sup> For example, by default the Google AdSense code that publishers insert into their webpages is asynchronous, meaning that the surrounding webpage can load before the ads do. (See Google, ‘An async script for AdSense tagging’, available [here](#), and also Google AdSense Help, ‘How to generate synchronous ad code for your ad units’, available [here](#).)

<sup>256</sup> One estimate based on scans of popular news websites in Europe found that websites made 68 calls per page taking 12 seconds on average. (ID5, User Sync Report, available [here](#)).

200. In principle, adtech providers should limit the number of cookie match requests that they make for users who already have a recent entry in the match table. Google, for instance, states that buyers should only serve the match tag if they do not already have a match for the user, or if the entry in the match table is older than 14 days.<sup>257</sup> Other adtech intermediaries have told us that the frequency depends on the partner, but that they match once per day as default.
201. A traditional cookie match shares no additional data beyond the cookie IDs, but it is possible to include more information (eg as extra URL parameters or headers on the cookie match requests). Once a cookie match has been made, most other data about the user can be transferred ‘offline’ at pre-set times (eg daily) using large batch files with user IDs and information for each user. Batch transfers are used because there are limits to the amount of information that can be transferred in real-time, and most information is not time-sensitive (eg demographic data). Offline data sharing is discussed more fully in the sections below on ‘Data management platforms and data brokers’ and use of tracking technologies and data for advertising in Google and Facebook.

### *Match rates*

202. Cookie matching is an imperfect process. At a basic level, cookies are associated to browsers rather than directly with people, so the activity of multiple users may be associated with the same cookie ID (eg a browser or device shared by multiple people), and multiple cookie IDs could exist for the same user (eg a person with multiple devices). In general, the use of cookies and other imperfect proxies for people can interfere with targeting, measurement, attribution and evaluation of digital advertising.<sup>258</sup>
203. We have heard that cookie match rates vary greatly, depending on which entities are doing the match, with a match rate 60% or above being considered ‘decent’.<sup>259</sup> A vertically integrated adtech platform told us that it considered the fact that it can offer advertisers 100% match rates between SSP and DSP a source of competitive advantage, which increases scale and improves performance when using audience data for targeting and frequency

---

<sup>257</sup> Google's Authorized Buyers Real-Time Bidding documentation on Cookie Matching, section on Cap frequency, available [here](#).

<sup>258</sup> For example, Facebook submitted research that it conducted in 2016 showing that using imperfect proxies for people like cookies could result in biased estimates of the effects of advertising. Coey, D., & Bailey, M. (2016, April). People and cookies: Imperfect treatment assignment in online experiments. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 1103-1111), available [here](#).

<sup>259</sup> See, for instance, Clearcode, What is Cookie Syncing and How Does it Work?, available [here](#).

capping. It stated that using a third-party ad exchange could result in an approximately 30% decrease in match rate.

204. More generally, if an advertiser must match data from several different sources, it will probably get an incomplete picture of users. By contrast, large incumbent platforms like Google and Facebook have a high reach, covering many people with their consumer-facing services, and could represent a single unified source of high-quality data, which improves the odds of successful matches and a higher match rate with the advertisers' own first-party data.

#### *Syncing mobile advertising IDs (MAIDs) is unnecessary*

205. Cookies and the need for cookie matching apply to web browsing activity on mobile devices. However, it is not necessary to use an equivalent matching or syncing process for mobile app activity; as discussed above, a MAID is unique to a mobile device and this MAID is shared with all native apps, enabling all app publishers to have a common identifier for each mobile device.
206. However, it may still be necessary to link MAIDs with cookie IDs to create cross-device graphs linking together all the different devices of a person, for example, using an IP address (discussed above in the section on 'Linking identifiers, identity resolution and cross-device tracking'), or first-party login details/internal IDs.

#### ***Data management platforms and data brokers***

207. This section focuses on the role of data brokers and data management platforms (DMPs) in the adtech ecosystem. These providers offer a set of related services, and some provide several of them, but the basic idea behind all of them is to collect, transact, store and manage data about people, for the purposes of targeting advertising and estimating the effectiveness of advertising. This includes facilitating cross-site tracking and linking data about a person across multiple contexts and properties owned by different data controllers.

#### *Data brokers*

208. Data brokers perform the following activities.
- (a) They collect, buy, or otherwise get access to data (including personal data) about people from a wide range of sources, including:

- (i) public sources (such as public registers);<sup>260</sup>
  - (ii) commercial sources (such as data about website visits and app usage from advertisers and publishers, but also data from market research firms and from other data brokers); and
  - (iii) primary research (such as panel surveys conducted directly on some consumers).
- (b) They analyse and combine data to make inferences about people, including about their demographics and interests, to create lists (or ‘audiences’, in the context of advertising) of people and identifiers. These data and inferences could include special category data, such as data about health, sexuality, racial or ethnic origin, and political and religious beliefs.
- (c) Data brokers provide these data, inferences, and audiences for a variety of purposes, including for digital advertising. These data, inferences and audiences are typically licensed to advertisers, media agencies, DSPs, DMPs and other data brokers, at a fixed cost or on a revenue share basis with the data supplier.
- (d) Data brokers help data suppliers to sell their data and audiences to others. Some entities provide a ‘marketplace’ where marketers can find and contact data suppliers, and purchase or import data directly from those suppliers. The marketplace platform may take a fee from third-party data providers, which is typically a proportion of the supplier’s revenue from data sales on the platform.
209. In common with most adtech providers that don’t have a direct relationship with consumers, typically data brokers rely on contractual requirements with their customers and partners to have a valid legal basis to process data about people (ie partners agree to obtain consent, in many areas of adtech including RTB). Where consent is relied upon, in theory, websites, apps and online services will have obtained valid consent from end-users before collecting and sharing data with data brokers and other adtech providers. (See also a related discussion in the previous section on ‘Transparency and Consent Framework (TCF)’.)
210. Notwithstanding efforts to provide notice and obtain consent, consumers are often unaware of the existence of data brokers. Because of this, there have long been concerns that consumers may not be able to exercise effective

---

<sup>260</sup> For instance, the open electoral register, register of company directors, HM Land Registry, the census and various area-level national statistics (which may be used to make probabilistic inferences about a person with a known address or postcode).

control of their personal data. For example, in 2014, the US FTC concluded that ‘to the extent data brokers offer consumers choices about their data, the choices are largely invisible and incomplete... because data brokers are not consumer-facing, consumers may not know where to go to exercise any choices that may be offered’.<sup>261</sup>

211. Although the ICO maintains a register of every organisation in the UK that processes personal data (unless they are exempt), the register does not include information on the purposes and data processed.<sup>262</sup> The most extensive lists of data brokers that we are aware are due to laws passed in February 2019 and October 2019 by the US states of Vermont<sup>263</sup> and California<sup>264</sup> respectively, which requires data brokers to register.<sup>265</sup> Although these registers are limited to data brokers that collect data on people in those states, in practice data brokers often hold data on people from around the world, including UK consumers.
212. It seems likely that data brokers collect and store information on a substantial proportion of UK people and households, although we have not directly investigated the extent of their coverage in the UK. For example, a recent study by Venkatadri et al. (2019) found that 74.4% of targetable Facebook identities in the UK are linked to data broker information in the UK, although we note that the population of targetable identities could be an unrepresentative sample of both the overall population of UK Facebook users (as some Facebook users may not be targetable due to their privacy settings) and of the overall UK population (as not everyone has a Facebook account).<sup>266</sup>

### *Data management platforms (DMPs)*

213. Data management platforms (DMPs) help advertisers and DSPs to import, manage, and ‘enrich’ data about people, and send that data back to DSPs and other parts of the adtech ecosystem (such as measurement and attribution providers) so that it can be used for targeting and other activities that make advertising more efficient.

---

<sup>261</sup> Federal Trade Commission, Data Brokers – A Call for Transparency and Accountability, May 2014, available [here](#).

<sup>262</sup> ICO, [Register of fee payers](#).

<sup>263</sup> Melendez (2019) ‘A landmark Vermont law nudges over 120 data brokers out of the shadows’, available [here](#).

<sup>264</sup> Office of the Attorney General of California, [Data Broker Registration](#).

<sup>265</sup> The California data broker registry is available [here](#). The data brokers registered in Vermont are discussed in Melendez and Pasternack (2019), ‘Here are the data brokers quietly buying and selling your personal information’, available [here](#).

<sup>266</sup> Venkatadri, G., Sapiezynski, P., Redmiles, E. M., Mislove, A., Goga, O., Mazurek, M., & Gummadi, K. P. (2019, May). Auditing Offline Data Brokers via Facebook's Advertising Platform. In The World Wide Web Conference (pp. 1920-1930). Available [here](#).

214. This includes advertisers' and publishers' first-party data, but also could be combined with second- and third-party data from other providers (such as data brokers) that the DMP partners with, provided that there is common identifier for users (such as matched cookies, or some other identifier like MAIDs or email address) between the advertiser and third-party data providers.
215. As discussed in the section on cookie matching, DMPs often act as a hub to sync IDs across the adtech ecosystem and host match tables which are shared across or available for all their customers, which helps to reduce the need for each pair of adtech participants to match with each other. Some DMPs specialise in 'identity resolution' (discussed in more detail in the section above on linking identifiers and identity resolution), to create profiles or identity graphs of individuals linking IDs across multiple adtech systems and all the identifiers from individual's multiple browsers, devices and different contexts.
216. DMPs offer tags and SDKs for advertisers to place on their websites and apps, which enable those DMPs to collect data on website visits and app usage directly on the advertisers' and publishers' behalf. Other data may be uploaded to DMPs by advertisers, including CRM records and transactions data from an 'offline' context (eg the records of a consumer buying something in a physical store against a loyalty card), and linked to online identifiers and profiles (eg the cookie ID for that consumer's browser when she logs into the store's website with loyalty card account, which was recognised when that consumer viewed an ad for that store on a different website).
217. Using these data, and like data brokers, DMPs also create and combine audiences (lists of people and IDs) focused on various targeting features such as demographics, interests (including being 'in-market'), and many others.
218. Depending on the provenance of the data inputs, and the specific data governance arrangements agreed with the relevant data suppliers, DMPs may restrict access to data and audiences only to those advertisers and publishers that supplied relevant inputs, or they can make these data and audiences widely available to DSPs and adtech providers that they work with. These processes are governed by data sharing agreements and could involve payment usually on either a fixed fee or a revenue share basis.
219. Google and Facebook both offer a similar function to DMPs by allowing advertisers to upload their data and lists (first-party data) to their respective platforms and use this for targeting advertising. These features – Google's Customer Match and Facebook's Custom Audiences – are discussed in their respective sections below.



## ***Use of tracking technologies and data for advertising in Google***

220. This section gives a broad overview of the main ways in which Google collects and sends data about users from and to third parties, and how this data is used for targeting advertising, measurement, and tracking conversions.
221. The scope of Google's collection of data on users for advertising-related purposes is vast. Google's Privacy Policy, which covers all of its consumer-facing services and products, but also 'products integrated into third-party apps and sites, like ads',<sup>267</sup> allows it to combine user data across all these contexts to linked identifiers and profiles of individuals.<sup>268</sup>
222. In general, Google's advertiser-facing products (eg Google Ads, DV360, SA360, Campaign Manager and Authorized Buyers) automatically collect certain user data when its advertising servers receive a request from a user's device (General User Data). This request may be triggered by the user interacting with a Google advertising service or with a third-party website or app that uses a Google advertising service. Depending on the publisher's settings, the user's preferences, and the device in question, General User Data may include:
- (a) the request itself;
  - (b) system and device information, such as the device, browser version, OS version, default language and screen size;
  - (c) IP address;
  - (d) GPS location – mobile devices with GPS functionality can, subject to user permissions, provide more granular location data than devices without it;
  - (e) date and time;
  - (f) for web browsers, the full URL of the page being visited together with the referrer URL;

---

<sup>267</sup> Google Privacy Policy, Introduction, available [here](#).

<sup>268</sup> Notwithstanding the general position in Google's Privacy Policy which allows it to collect and combine user data across all its services for various purposes including delivering personalised advertising, Google told us about two exceptions: 1) Gmail – Google stated that it does not use Gmail data to personalise advertising. The ads that are shown in Gmail are completely independent from the user's content within the Gmail service (ads are selected based on users' general profile); and 2) Google Sign-In – Google stated that it does not collect data from third-party sites and apps and services via Google Sign-In about the user's activity in that app. Google Sign-In does store the context under which the user authenticates, like information about the device that was used, IP address, and identifiers for the app to which the user has authenticated. This helps prevent abuse and provide transparency and control to users about which apps they can sign into via Google Sign-In, and allows them to revoke access.

- (g) for mobile devices, mobile network information;
- (h) for mobile apps, an identifier for the app and the MAID;
- (i) for web browsers, any cookie IDs that Google has previously set on the user's device; and
- (j) event data – such as impressions, clicks or conversions.

223. Google may collect different data depending on whether the user is signed-in to their Google Account.

- (a) Google stores data against the user's Google Account if the user is signed in and has consented to this.<sup>269</sup> Subject to these user permissions, Google may combine activity from the user across devices and browsers, including for the purposes of ad targeting and measurement.
- (b) If the user is not signed into a Google Account, Google will store data against a unique resettable identifier linked to the user's browser or device, such as a cookie ID or a MAID.

### *Publishers*

224. Through its publisher-facing products (AdSense, AdMob, and Google Ad Manager), Google collects General User Data through tags on publishers' properties and Google Mobile Ads SDK on users' apps. These tags and SDKs are used to create ad requests to Google's advertising servers.

- (a) For websites using AdSense, Google displays ads using an iFrame. When the user's browser renders the webpage with the ads, the iFrame will direct the browser to Google's servers. Google places a cookie on the user's device to record information, such as the user's interaction with Google's ads.
- (b) For AdMob apps, Google may collect data using SDKs and MAIDs, as governed by Google's AdMob terms of service.

225. Google Ad Manager also allows publishers to upload their user IDs for use in frequency capping, audience segmentation, and targeting, amongst other purposes.<sup>270</sup> Google Ad Manager also allows publishers to integrate audience data, such as audience lists and lists of cookie IDs with inferred interests, from their own DMPs, and these audiences can be offered by the publisher to

---

<sup>269</sup> Google provides users with granular information and privacy controls for Google Accounts. (More on this in the section on User Control and Tracking, and in Appendix K.)

<sup>270</sup> Google, About publisher provided identifiers, available [here](#).

advertisers for publishers to target on the advertisers' behalf during campaign negotiations.<sup>271</sup>

226. For completeness, Google also collects user data from website operators and app developers from Google Analytics. Google provides website operators with analytical data on user interaction with their website via Google Analytics. (This is discussed in more detail on in the section below on Google Analytics.)
227. Google provides reporting back to website operators and app developers. In the case of Google Ad Manager, event level data can be provided to a customer through Data Transfer report files. (These are discussed in more detail in Appendix M and later section on 'Interactions between market power and tracking'.)

## *Advertisers*

### *Ad serving*

228. Through its advertiser-facing products (Google Ads, DV360, SA360, Campaign Manager, and Authorized Buyers), Google collects General User Data when serving an ad – this could be on a Google property and on a partner publisher's property (third-party inventory in AdSense and AdMob networks).
229. As discussed above, in the section on Programmatic display advertising and real-time bidding, Google's ad exchange (Google Ad Manager/Authorized Buyers) sends a subset of General User Data in bid requests using the applicable RTB protocol. Google's DSPs (DV360 and Google Ads) receive data in bid requests from ad exchanges (both third-party exchanges and Google Authorized Buyers).
230. DV360, Google Ads, and Campaign Manager may also receive publisher click tracker URLs, publisher-defined values, advertiser-defined values and other values and identifiers.

### *Cookie matching and use of DMP data for targeting*

231. In addition, advertisers can use the Authorized Buyers Bulk Uploader API to create and edit user lists of cookie IDs and MAIDs to store with Google (for pre-targeting). As discussed in the section on cookie matching, buyers can use the Cookie Match API to map cookie IDs in their domain to cookie IDs in the Google domain. This mapping may be hosted by Google or by the buyer

---

<sup>271</sup> Google, Introduction to Audience Solutions, available [here](#).

themselves and may be used by the buyer to bid efficiently against other demand sources.

232. In addition, advertisers can integrate DV360 audience data (such as audience lists and lists of cookie IDs with inferred interests) from their own data management platforms and third-party vendors.

#### *Remarketing lists and Customer Match*

233. Google allows advertisers to upload or provide their own data by creating remarketing lists, including through the Customer Match feature. They can use these remarketing lists to target users on those lists directly. Google does not share remarketing lists with any third-party and other advertisers without permission.<sup>272</sup> Google can also identify users with matching behaviour to find similar audiences. Google provides reporting data to advertisers and publishers, which generally aggregated and anonymised.<sup>273</sup>
234. Google Ads and DV360 (for TrueView line items)<sup>274</sup> allow advertisers and authorised third parties (such as media agencies and DMPs acting for those advertisers) to upload data about their customers through a feature called Customer Match.<sup>275</sup> This data includes names, email addresses, mailing addresses and phone numbers that advertisers collected directly from users.
235. Google uses uploaded data to Customer Match to match the advertiser's customers with Google users, subject to users' settings for ad personalisation on their Google Account, to enable advertisers to create audience lists and to target, bid differently for and show customised ads to users on those lists.
236. There are also other optional Google Ads features that allow advertisers to upload additional user data, such as MAIDs, for the purposes of ad measurement or ad targeting.
237. The contractual provisions governing conversion data and customer matching are the Google Ads terms of service together with applicable policies,

---

<sup>272</sup> See Google, How Google uses remarketing data, available [here](#).

<sup>273</sup> We note that the extent to which anonymisation is possible is an open research question. The susceptibility of anonymised data to re-identification or de-anonymisation depends on context and capabilities of potential attackers. We have not examined these possibilities in detail with respect to the reporting data that Google sends to third parties.

<sup>274</sup> Customer Match is only available via DV360 for TrueView line items. See Google, Customer Match for TrueView line items, available [here](#). (TrueView is a cost-per-view, choice-based video ad format that Google offers on YouTube, apps and websites, implementing various conditions so that a view is counted, and advertisers pay only when the user is deemed to have actually viewed the ad. See Google, About TrueView line items, available [here](#).)

<sup>275</sup> See Google, About Customer Match, available [here](#).

including Google's policies on Data Use in personalised ads, and its policy on use of Customer Match data. For example:

- (a) Google restricts advertisers' use of third-party data partners for new customer prospecting on YouTube, Search, and Gmail. For these products, advertisers can use their own customer lists, use Google's user list and associated data, and use purchased/obtained data to further segment customer lists, but cannot use purchased/obtained customer lists directly.<sup>276</sup>
- (b) Google agrees to use data collected via Customer Match only for the advertiser that uploaded the data.<sup>277</sup>

### *Conversion tracking and Store Sales*

- 238. Google Ads also collects General User Data when tracking conversion on advertisers' properties. Advertisers can use Google Ads to measure the effectiveness of their ads through conversion tracking (if advertisers choose to share certain data with Google Ads for this purpose), by placing a tag or pixel on their websites or including an SDK on their apps to record online conversion events (such as website/app purchases, newsletter signups, button clicks and app installations).
- 239. Google may also receive data about offline conversion events (such as phone calls and store visits). Similar to Customer Match, Google Ads also allows advertisers and authorised third parties (such as media agencies and DMPs acting for those advertisers) to upload data about transactions through a feature called Store Sales.<sup>278</sup> Again, this data could include names, email addresses, mailing addresses and phone numbers that advertisers collected directly from users. Google Ads uses uploaded data to match the advertiser's customers with Google users, subject to users' settings for ad personalisation on their Google Account, to measure offline conversions.<sup>279</sup>
- 240. Google's policy on its use of conversion event data states that Google doesn't share advertiser-specific conversion event data with other advertisers (unless

---

<sup>276</sup> Google, Data use in personalised ads on Google Search, Gmail, and YouTube, available [here](#).

<sup>277</sup> Google, How Google uses Customer Match data, available [here](#) (and an equivalent page for DV360 and Google Marketing Platform [here](#)).

<sup>278</sup> See Google, About store sales conversions (available [here](#)) and related links. Store Sales is not available via DV360. Whilst Store Sales itself does not provide ads personalisation, advertisers can use the same data they uploaded for Store Sales for Customer Match.

<sup>279</sup> Google provides a number of other offline conversion tracking methods beyond store sales. These include phone call conversions (which requires the use of Google forwarding numbers) and shop visit conversions (which uses data on ad exposures, and anonymous and aggregated phone location history for users that have consented to share their location history). See Google, About offline conversion tracking ([here](#)), and related pages.

it has the advertiser's permission). However, we note that Google's policy on its use of conversion event data also explicitly allows for uses of aggregated conversion event data for the overall benefit of advertisers. For example, features such as automated bidding and smart pricing rely on aggregate advertiser conversion event data to improve their overall quality and accuracy.<sup>280</sup>

### *Reporting*

241. Google provides its advertising partners with reporting data so that those partners can evaluate the performance of Google's ads and optimise their bidding strategies. Reporting data is generally aggregated and anonymised.<sup>281</sup> Provision of such data occurs under Google's template terms of service for its advertiser-facing services.

### *Data brokers and other adtech providers*

242. Google told us that the main circumstances and contractual arrangements under which other entities provide user data to Google, and where the data exchanged may be used for advertising purposes by either Google or the other entity, are largely limited to data shared with advertisers and publishers in the context of providing services to them. Google stated that it does not sell or otherwise provide user data to third parties in return for consideration.

### *Data brokers*

243. Google may obtain data from data brokers, but only to market its own products and services. In general, agreements with data brokers typically provide for non-exclusive data licenses to Google, with the data broker retaining ownership over the data provided. Google stated that it does not use such data within its advertising services or to build user profiles.
244. Google further stated that it does not combine data for purposes of identification or otherwise engage in fingerprinting, apart from actions to prevent fraud or abuse of Google's products.

---

<sup>280</sup> Google, How Google uses conversion event data, available [here](#).

<sup>281</sup> We note that the extent to which anonymisation is possible is an open research question. The susceptibility of anonymised data to re-identification or de-anonymisation depends on context and capabilities of potential attackers. We have not examined these possibilities in detail with respect to the reporting data that Google sends to third parties.

### *Google Measurement Partners program*

245. Google allows advertisers to measure whether an ad campaign resulted in an increase in sales or brand awareness through tools that help measure the effectiveness of particular ad types. For example, Google offers tools such as the Brand Lift tool to help advertisers measure the effectiveness of their video ads.<sup>282</sup> Another set of tools available to advertisers for measuring the effectiveness of video ads is delivered via the Google Measurement Partners program.
246. The Google Measurement Partners program involves more than 20 verified partners (such as Nielsen and Sizmek) to help advertisers measure the effectiveness of video ads across seven specialisations: viewability, reach, brand safety, brand lift, sales lift, app attribution, and marketing mix modelling. Google provides data to these third-party partners, who offer solutions that work across Google advertising products, including Google Marketing Platform, Google Ads, and YouTube. Google provides customer data to such partners under Customer Data Sharing Agreements.

### *Google Analytics, Floodlight, and Google Tag Manager*

247. This section sets out a broad overview of Google's analytics services, and their interaction with Google's advertising services.
248. Google Analytics is used by website or app owners to track site or app activity such as session duration, pages per session and bounce rate of individuals visiting the site or app, along with information on the source of the traffic. Analytics customers add the Analytics tracking code to their website and/or the Firebase SDK in their app, which collects data from the customer's website/app, and returns that data to Google Analytics where the customer can see it in reports and understand their properties' performance.
249. Floodlight is a conversion tracking and reporting system that advertisers and media agencies use through DV360, SA360 or Campaign Manager. It uses tags and cookies to track conversions from display and search advertising. To achieve this, it is necessary to work out whether an ad exposure and subsequent conversion event involved the same user. A Floodlight tag is an iFrame or image tag that advertisers can install on a relevant conversion page on their site. The information recorded by the Floodlight tag depends on the configuration by the advertisers and the location on the advertiser's website. When a user lands on the conversion page that contains a Floodlight tag, the

---

<sup>282</sup> The Brand Lift tool works through consumer surveys (such as the one question surveys on YouTube videos) and measuring changes in search volumes in response to ad campaigns (eg comparing organic search behaviour of treatment and control groups).



tag sends data about the page view to DV360, SA360 and/or Campaign Manager. As part of this process, Campaign Manager checks the user's DoubleClick cookie to see whether the user has previously viewed or clicked on the advertiser's ad. In this case, the page view is counted as a view-through conversion.

250. Google Tag Manager is used by website or app owners to manage their tags and codes through a single container that can replace other manually coded tags on a site or app, including tags from Google Ads, Google Analytics and Floodlight. Google Tag Manager may collect some aggregated data about when tags are activated, in order to monitor system stability and performance. This data does not include user IP addresses or any user-specific identifiers that could be associated with a particular individual. Other than data in standard HTTP request logs (all of which is deleted within 14 days of being received) Google Tag Manager otherwise does not collect, retain, or share any information about visitors to its customers' properties, including page URLs visited. Apart from this, Google Tag Manager does not itself collect data. For this reason, Google Tag Manager is not discussed any further in this section.

#### *Interaction between Google's analytics services and advertising services*

251. Google told us that it only uses data from Google Analytics for its own purposes if the customer has enabled data sharing with Google. With regards to Floodlight, Google does not use data collected by Floodlight data for its own purposes. All use of Floodlight data is directed by the advertiser.<sup>283</sup>
- (a) Google Analytics' terms of service do not allow customers to send Google personally identifiable data, so this data is not used for ads personalisation.<sup>284</sup> Google Analytics customers must also not circumvent any privacy features that are part of Google Analytics.<sup>285</sup>
- (b) The collection and processing of data by Google Analytics is governed by the Google Analytics terms of service. Google cannot use data collected from a website via Google Analytics for purposes other than providing

---

<sup>283</sup> In addition, Google stated that it has strict internal controls on access to data collected via Google Analytics. For example, Google engineers are only permitted to access Analytics data where they can demonstrate a Google Analytics business need to do so and are only able to access the data which pertain to their specific pipeline. Individuals who are provided access to Google Analytics account data, including vendors, must agree to internal access policy terms and conditions, must obtain appropriate authentication and must use Google-approved computers. [§<

<sup>284</sup> Google Analytics Terms of Service, section 7 on Privacy, available [here](#).

<sup>285</sup> Google Analytics' terms of service also prohibit customers using Google Analytics Advertising Features (discussed below) from facilitating the merging of personally-identifiable information with non-personally identifiable information collected through any Google advertising product or feature unless they have robust notice of, and the user's prior affirmative (ie opt-in) consent to, that merger.

analytics services to that website's operator, unless the operator has specifically agreed otherwise.<sup>286</sup>

- (c) Google only uses data from Google Analytics for its own purposes if the customer has not opted out of the 'share data with Google products and services' setting.<sup>287</sup> Google told us that this data sharing is currently only enabled for approximately [20-30%] of Google Analytics accounts. If customers have enabled data sharing with Google, Google uses Google Analytics data for various purposes, including benchmarking, support, sales, and improving Google's products (though the exact uses will depend on which specific data sharing functionalities the customer has enabled).<sup>288</sup>
- (d) Google further stated that, regardless of the data sharing settings, Google Analytics data may also be used only insofar as necessary to maintain and protect the Google Analytics service (ie for the detection and prevention of misuse, abuse, spam, malware etc).
- (e) Similarly, Google stated that it does not use data collected by Floodlight for its own purposes. Floodlight data and derivatives (such as remarketing lists based on Floodlight data) is owned by the advertiser, and all use of the data is directed by the advertiser.<sup>289</sup>

252. Instead, Google told us that advertisers and publishers may use Google Analytics insights to improve the performance of their ads and properties. For example:

- (a) A website publisher can link other Google services that it uses, eg Google Ads and AdSense, with Google Analytics to obtain reports on these other Google services.<sup>290</sup>
- (b) Advertisers can create remarketing lists in Google Analytics based on user behaviour on their sites (for example based on users who purchased

---

<sup>286</sup> Google further stated that it would not use data from third-party partner sites such as remarketing pings to determine a user's interests.

<sup>287</sup> Google stated that this is an affirmative choice, and customers can later decide to disable the data sharing setting at any time even if they set up their Google Analytics account with data sharing enabled. Customers do not receive a reduced Google Analytics service for choosing not to share their data with Google.

<sup>288</sup> Google Analytics Help, Data Sharing Settings (available [here](#)), states: "There are several data sharing settings in your Analytics account... Regardless of your data sharing settings, your Analytics data may also be used only insofar as necessary to maintain and protect the Analytics service". One of these settings is 'share data with Google products and services'. In a later section titled 'Details and benefits of each data sharing setting', the help page states 'When you turn [the 'share data with Google products and services'] setting ON, Google can access and analyze data to better understand online behavior and trends, and use this data to improve Google products and services. For example, this data can be used to improve the Google Ads system tools that you use to create, manage, and analyze your ad campaigns.'

<sup>289</sup> For instance, the advertiser may enable related entities (eg affiliated companies) to use the same remarketing lists chosen and created by it, which consists of cookies associated with browsers that have visited that advertiser's website.

<sup>290</sup> Google Analytics Help, Link/unlink Google Ads and Analytics, available [here](#).

specific items, or users who looked at certain items, but chose not to make a purchase), and then target those audiences for remarketing campaigns via ad accounts such as Google Ads or DV360.

- (c) Google Analytics can collect user-level data and aggregate it into reports for the customer to enable that customer to measure the effectiveness of their ads. If the customer chooses to integrate their various ad products, they can use Google Analytics reports to make changes to various other functionalities in those other accounts. For example, while Google Analytics is not directly involved in making changes to bidding, advertisers can use the insights from Google Analytics to make better bidding decisions.
- (d) If Google Analytics customers enable Analytics Advertising Features,<sup>291</sup> Google Analytics can also collect information about the customer's users from any Google advertising cookies and identifiers when they are present. Google Analytics advertising reporting features include Google Display Network impression reporting and Google Analytics demographics and interest reporting. The reporting features on Google Analytics are based on the data shared by the Google Analytics customer.
- (e) Advertisers can integrate Floodlight with DV360 audience data (such as audience lists, lists of cookie IDs with inferred interests) from their own DMP and third-party vendors.

### *Interoperability*

- 253. Google Analytics can be used separately or together with other services in the Google Marketing Platform. It doesn't distinguish between ads intermediated by third parties or ads intermediated by Google. Google Analytics can be integrated with third-party services, such as ad networks, ad servers, DSPs and SSPs, and search engine management tools, using the Google Analytics data import feature.
- 254. We have not directly assessed whether Google Analytics works better with other services in the Google Marketing Platform relative to competing adtech services.

### *Consent*

- 255. In common with most adtech providers processing data on behalf of third-parties that hold the direct relationship with end-users, Google relies on contractual terms that require customers to disclose (eg in privacy policies) to

---

<sup>291</sup> Google, About Advertising Features, available [here](#).

end-users that their website or app is using Google Analytics or Google Tag Manager and how these products collect and process data, and where required by law, obtain end user consent to the storage and access of cookies in connection with their use of Google Analytics.<sup>292</sup>

256. Google told us that for compliance purposes it conducts manual reviews of websites and apps that use Google advertising services, and [X]. Google's reviewers visit a site or app as a consumer would visit, and look at the information provided and the consents obtained.

### ***Use of tracking technologies and data for advertising in Facebook***

257. This section gives a broad overview of the main ways in which Facebook collects and sends data about users from and to third parties, and how this data is used for targeting advertising, measurement, and tracking conversions. It discusses Facebook's Business Tools and Custom Audiences.

258. As discussed in Appendices F and O, Facebook allows advertisers to i) target users based on demographics, interests and location and ii) track conversions, using information volunteered, observed and inferred from:

- (a) users' activity on Facebook Products (such as Facebook, Messenger, and Instagram); and
- (b) users' 'off-Facebook activity', such as information shared with Facebook by websites and apps that send Facebook data directly using Facebook Business Tools (also discussed below in the section on Facebook Business Tools) and by advertisers who upload customers lists (discussed below in the section on Facebook Custom Audiences).<sup>293</sup>

### ***Facebook Business Tools***

259. The Facebook Business Tools are a number of products and services that Facebook offers to enable website owners and publishers, developers, advertisers, business partners (and their customers) and others to integrate, use and exchange information with Facebook, subject to users' browsers and device settings.

260. The main Facebook Business Tools through which Facebook may receive data from third parties that is used for its advertising services are: (i) the

---

<sup>292</sup> Users can opt out of data collection by Google Analytics from sites accessed using a browser by downloading and installing the opt-out add-on for their browsers. However, as discussed in the section on 'Fingerprinting', installing add-ons may increase users' vulnerability to browser fingerprinting, as the presence of the add-on increases entropy.

<sup>293</sup> Facebook, [About Facebook Ads](#).

Facebook Pixel; (ii) Facebook SDKs; (iii) Facebook Login; and (iv) social plugins.

261. Facebook uses cookies to enable it to offer various services and features (eg the Facebook Pixel) and to understand the information third parties choose to share with Facebook.
262. Facebook pools the data about users that it obtains from advertisers using Facebook Business Tools and its own products, and it uses this data to personalise ads and recommendations.<sup>294</sup>

### *Facebook Pixel*

263. Facebook Pixel is a small piece of code that business customers (mainly advertisers) can choose to add to their websites, to build audiences and to measure and refine ad campaigns. When a user visits a website that has the Pixel, the Pixel is triggered and Facebook's servers automatically log:
  - (a) the fact that a particular browser visited the website;
  - (b) HTTP headers (IP address, info about the web browser, page location, document, referrer and person using the website);
  - (c) Pixel-specific data – Pixel ID, and data used to connect the events to a specific Facebook ad account and to make a match to a person known by Facebook (eg Facebook cookie, timestamp, Pixel version, possibly nature of event such as 'add to cart' or 'purchase'); and
  - (d) page metadata.
264. Additionally, websites that have implemented the Pixel may choose to share the following additional information with Facebook:
  - (a) Button click data. This includes any buttons clicked by the person on the website, the labels of those buttons and any pages visited as a result of the button clicks.

---

<sup>294</sup> All of Facebook Business Tools are covered by a single set of terms, which state that: "We use Event Data [information that advertisers share about users' actions on the advertisers' properties] to personalise the features and content (including ads and recommendations) that we show people on and off our Facebook Company Products. In connection with ad targeting and delivery optimisation, we will: (i) use your Event Data for delivery optimisation only after aggregating such Event Data with other data collected from other advertisers or otherwise collected on Facebook Products; and (ii) not allow other advertisers or third parties to target advertising solely on the basis of your Event Data." Facebook Business Tool Terms are available [here](#).

- (b) Optional values. Third parties implementing Pixel can choose to share additional information about users' visits with Facebook, including through the use of custom events.
  - (c) Form field names. This includes website field names such as 'email', 'phone', 'name', 'address' and 'quantity' when a person purchases a product or service.
265. Facebook enables websites to delay the Pixel from firing until the website owner has obtained affirmative consent from the user.
266. Websites can choose to implement Facebook Pixel using both first-party and third-party cookies, and it uses both by default.<sup>295</sup> This means Facebook Pixel can work even when browsers are blocking third-party cookies.<sup>296</sup> (See also the section above on 'Third party code in first party websites'.)

### *Facebook SDK*

267. SDKs are bundles of code that offer extra functionality to a developer who incorporate the SDK into their app. For example, as part of the SDK infrastructure, Facebook is able to log API calls (eg when a user opens and closes an app), which allows Facebook to provide analytics back to the app developer on, for example, the usage of their app. Mobile apps or websites that integrate Facebook SDKs can enable Facebook securely to receive information about the actions of users on the app or website.<sup>297</sup>
268. When a user opens an app that has implemented the Facebook SDK, subject to users' browser and device settings, the app shares the following data with Facebook:
- (a) automatically logged events (basic interaction in the app, such as app installs and app launches);
  - (b) app info (SDK version, Facebook App ID, app name and version); and
  - (c) device-related info (MAID, OS version, time zone, country, language, model, User Agent string, IP address, mobile network, screen size, cores and disk space, device opt-out settings).
269. Depending on the app and user's browser and device settings, the app may share the following data with Facebook via Facebook's SDKs:

---

<sup>295</sup> Facebook, [About cookie settings for Facebook pixel](#).

<sup>296</sup> See, for instance, Clearcode, [What Facebook's First-Party Cookie Means for AdTech](#), available [here](#).

<sup>297</sup> See Facebook's SDKs for Android and iOS, and Facebook Business SDK on [Facebook for Developers APIs and SDKs](#).

- (a) explicit events: information from events that the customer configures their app or website to share, such as 'Add to Cart' or 'Purchase', along with any additional parameters provided. App and websites can customise the event data they share with Facebook in a number of ways, including through the use of custom events; and
  - (b) implicit events: metadata relating to information from events, such as interactions with Facebook Login or the 'Like' button that are logged implicitly along with automatic or explicit events or if the customer chooses to make use of other features of the Facebook SDK.
270. Facebook does not receive any data if a user downloads but does not open an app that has implemented the Facebook SDKs for iOS or Android. Facebook only receives data that the app chooses to share with it through the Facebook SDKs for Android or iOS after a user has opened the relevant app.
271. Facebook's SDKs for Android and iOS enable the app developer to disable the transmission of the above information (such as app installs and app launches) to Facebook and/or delay it until after the user has completed the developer's in-app consent flow and thereby consented to this information being shared.
272. Facebook Login and social plugins (discussed below) can be implemented through Facebook SDKs.

### *Facebook Login*

273. Facebook Login can be integrated in websites and apps to enable people to use their Facebook account to log in to the relevant website / app. Customers can incorporate Login into their website using the Facebook JavaScript SDK<sup>298</sup> or by downloading the relevant code from Facebook's website and including it on their website by adding an element to their code (ie by building a manual login flow).<sup>299</sup>

### *Facebook social plugins*

274. Social plugins, such as the 'Like' and 'Share' buttons, can be incorporated in off-Facebook websites and apps to enable them to provide social experiences.<sup>300</sup>

---

<sup>298</sup> Available here: [Facebook for Developers APIs and SDKs](#).

<sup>299</sup> Facebook for Developers, Manually Build a Login Flow, available [here](#).

<sup>300</sup> The social plugins currently offered by Facebook include: (i) Comments; (ii) Embedded Comments; (iii) Embedded Posts; (iv) Embedded Videos; (v) Group Plugin; (vi) Like Button; (vii) Page Plugin; (viii) Quote Plugin; (ix) Save Button; and (x) Share Button. For a complete list of the Facebook Social Plugins and each plugin's program code, see [here](#).



275. Websites and apps can incorporate social plugins in a similar way to Facebook Login (eg using the Facebook SDKs for Android and iOS).

### *Facebook Custom Audience and Offline Conversions*

276. Facebook offers a feature called Custom Audience, which allows advertisers to target advertising on Facebook to their existing (or prospective) customers. One of the ways in which an advertiser can create a Custom Audience is by providing Facebook with a hashed list of identifiers (such as emails and phone numbers). Facebook applies the same hashing algorithm to its own set of user provided identifiers, including those provided at registration (but only for those users that have not used the Facebook settings to disconnect 'off-Facebook activity'), compares the two sets of hash values, and places matched users into a Custom Audience list available for that advertiser to use for targeting, conversion tracking, and also audience expansion (called Lookalike Audience on Facebook).
277. Hashing help maintains privacy somewhat because it protects against Facebook receiving the personal information of anyone for whom it does not have a match – eg an advertiser's customer that is not a Facebook user. However, it is important to note that a hashed identifier is nevertheless a persistent, unique identifier that allows linking a person across databases, devices, and contexts.
278. Facebook imposes a minimum size of Customer Audiences lists from advertisers.<sup>301</sup> After the match, Facebook reveals the approximate number of matches to the advertiser but not the identities of the matched users.<sup>302</sup> Facebook deletes all hashes (both matched and unmatched) within 48 hours after the matching, and no further processing of the hashed values is performed beyond the match process.
279. Facebook also allows advertisers to share information about offline conversions with Facebook using Facebook's Offline Conversion API, or through one of Facebook's partner integrations. This enables advertisers to

---

<sup>301</sup> According to [Facebook's documentation for its Customer Audience API](#), the minimum size of the origin audience is 100.

<sup>302</sup> According to Venkatadri et al. (2018), Facebook does not report any size statistics for audiences creates using multiple personally identifiable information (PII) attributes, and no size estimates when combining audiences that were created using different PII attributes. This was in response to the authors demonstrating that there were able to exploit previous vulnerabilities in Facebook's Custom Audience interface that allowed an attacker to infer users' full phone numbers from knowing just their email address, determine whether a particular user visited a website, and de-anonymise all the visitors to a website by inferring their phone numbers. Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., & Goga, O. (2018, May). Privacy risks with Facebook's pii-based targeting: Auditing a data Broker's advertising interface. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 89-107). IEEE, available [here](#).

identify correlations between Facebook ad campaigns and in-store (offline) purchases.<sup>303</sup>

#### *Data that Facebook provides to third parties*

280. Facebook stated that it does not provide data that directly and personally identifies a user, and that it only provides aggregated and anonymised data to third parties.<sup>304</sup>
281. Facebook does not give third parties access to the data received via Pixel and SDK for Android and iOS, other than to the third party that has installed the relevant Pixel or SDK or if instructed to do so by the installing third party.<sup>305</sup>
282. Facebook may share metrics and insights concerning users' activities to enable them to better target, manage and evaluate their advertising campaigns. However, the data shared in this context does not contain disaggregated data received via an SDK or Pixel (eg data on what a user purchased) and is appropriately anonymised and aggregated (including, for example, with data from other sources, such as by being combined with data relating to users' on-Facebook activities).
283. Some of the main ways that Facebook provides data to third parties are:
- (a) Facebook Ads Manager (and Business Manager, for customers with multiple ad accounts for separate campaigns) – Facebook provides aggregated metrics on how the customers' ads are performing along various dimensions (such as across different demographics, etc.).
  - (b) Facebook Audience Insights – Facebook provides aggregated and anonymised information on users including demographics, page likes, location and language, usage, purchases activity. This information is provided for Facebook users generally, and for users connected to Facebook's business customers' Pages or Events on Facebook, or for users in Custom Audiences they have created.
  - (c) Facebook Analytics – Facebook Analytics provides businesses with aggregate insights about users' interactions with their websites and apps (in addition to their interactions with their businesses' Facebook Pages),

---

<sup>303</sup> See Facebook Help Centre, [About offline conversions](#).

<sup>304</sup> We note that the extent to which anonymisation is possible is an open research question. The susceptibility of anonymised data to re-identification or de-anonymisation depends on context and capabilities of potential attackers. We have not examined these possibilities in detail with respect to the data that Facebook sends to third parties.

<sup>305</sup> For example, when an advertiser uses a third-party platform or a tag manager (eg Google Tag Manager) to implement a Pixel on its website.

which can be gathered if they place Facebook's code (eg Facebook's Pixel or SDKs) on their website or app, or work with one of Facebook's measurement partners (discussed below on Facebook Marketing Partners). Facebook provides aggregated insights about those users, such as demographic profiles, interests and types of devices used. Businesses use these insights to improve the relevance of their content and features to those users.<sup>306</sup>

- (d) Facebook Attribution – Facebook provides self-serve tools for advertisers (including Brand Lift and Conversion Lift, which are discussed in more detail in Appendices F and O), which enable advertisers to compare the relative effectiveness of their ads across Facebook's services and third-party publishers through measurement tags. Advertisers can then: (i) view the order in which a user sees an ad across channels (both on- and off-Facebook); and (ii) assess the relative impact of each channel, attributing credit for a conversion to the specific ad publishers that contributed to the user's journey to a conversion.<sup>307</sup>
- (e) Facebook Marketing Partners – Facebook partners with over 40 companies and other entities around the world who provide tools and independent metrics to advertisers for Facebook ads.<sup>308</sup> These metrics fall into five broad areas: reach, viewability, attribution, brand lift and outcome lift. Facebook sends data (Facebook used to work with some third-party data providers to offer their targeting segments, called Partner Categories, directly on its platform but has terminated this program in May 2018.<sup>309,310</sup> Facebook's advertiser customers can continue to work with third-party data providers, but they are required to have any necessary rights and permissions to use this information when using Custom Audiences.)
- (f) Facebook Platform – Facebook Platform enables third-party developers to interact programmatically with the Facebook Service, including the Graph API, and also partner-specific APIs for various purposes.

---

<sup>306</sup> See Facebook Analytics (available [here](#)), and Facebook's documentation for Facebook Analytics ([here](#)).

<sup>307</sup> Facebook Help Centre, Get started with Facebook Attribution, available [here](#).

<sup>308</sup> Facebook Marketing Partners, available [here](#).

<sup>309</sup> See Facebook, 'How does Facebook work with data providers?', available [here](#).

<sup>310</sup> Twitter has made a similar announcement in August 2019 that it will remove third-party data options from its ad targeting process. See Social Media Today, '[Twitter Announces Removal of Third-Party Data Sources from Ad Targeting Options](#)'. At the time of writing, Twitter still partners with a range of data providers (see Twitter Data Partners [here](#)).

## *Consent*

284. Facebook relies on the user's consent to process data received from third parties, such as websites and apps that use the Facebook Business Tools (ie off-Facebook activity), for purposes of targeting ads to that user.
285. Under the Business Tool Terms, Facebook requires third parties using the Facebook Business Tools to warrant that they provide clear and sufficiently prominent notice to inform users visiting their websites or apps about the data being shared with Facebook and where necessary to obtain users' valid consent.<sup>311</sup>

## **Estimates of the extent of tracking by Google, Facebook and other market participants**

286. Currently, tracking is necessary for many activities which enhance the efficiency of personalised advertising, such as targeting, measurement (including the detection of invalid traffic) and attribution.
287. Access to opportunities for tracking and data about users are a source of competitive advantage and market power. Estimates of the prevalence of tracking suggest that users' activities are tracked across most websites and mobile apps (including the most popular web properties), but also that large incumbent platforms such as Google and Facebook have greater opportunities to track and collect data on users than other advertisers, publishers and adtech providers.
288. This section:
- (a) recaps the incentives underlying third-party advertisers' and publishers' permitting of Google and Facebook to track their users on third-party properties; and
  - (b) presents an overview of the evidence available on the prevalence and prominence of tracking on websites and mobile apps.

## ***Market power, relative value of, and access to data and opportunities for tracking***

289. As explained in Appendix F, many advertisers and publishers use Google and Facebook's advertising services, in large part because of their strong position

---

<sup>311</sup> Facebook Business Tools Terms, available [here](#).

in their various consumer-facing services such as Search, YouTube, and Facebook, which:

- (a) makes their properties important sources of inventory on which to advertise to a large group of users, and;
  - (b) gives Google and Facebook a lot of insight into users which can be used to target advertising and improve the performance of ads that they deliver, both on their properties but also on third-party publishers' properties.
290. As discussed in Appendix F, Google is particularly privileged, from its position as the dominant general search engine, in its ability to obtain insights about whether users are 'in-market' and actively searching for products and services. The insight from these data are highly valuable to advertisers, and are not easily available to advertisers, publishers, data brokers, and other adtech providers from elsewhere. By contrast, one respondent told us that its data and audiences, and those of other similar businesses, are widely available and accessible across a broad range of media platforms. In its view, barriers to entry are relatively low for providers to create and offer its audiences for digital advertising purposes.
291. The difference in access and quality of data that large platforms like Google and Facebook have relative to other market participants affects competition in digital advertising. In general, it is difficult and costly for advertisers to assemble information on consumers, compared to Google/Facebook, from their own first-party data and other (non-Google/Facebook) third-party data providers. Google and Facebook have high reach, as many people use them, and Google in particular has very valuable data on consumers that can be used to target advertising. As set out in the previous sections on the Use of tracking technologies and data in Google and Facebook, Google and Facebook do not provide access to this data on open data exchanges, so the only way for advertisers to get (indirect) access to it and use it for targeting is to use Google and Facebook's ad management tools.
292. As a result of the importance of Google and Facebook's own properties as sources of inventory on which to advertise, many advertisers use Google and Facebook. Google and Facebook have limited interoperability by not providing user- and event-level data outside of their systems to third parties. Therefore, advertisers cannot frequency cap or measure user exposures consistently across all inventory using third-party measurement and attribution tools, but can do so with Google and Facebook's tools. This gives advertisers an incentive to just use Google and Facebook's tools. This is discussed in more detail in Appendices O and M.

293. Advertisers using Google and Facebook’s advertising services have an incentive to install Google and Facebook’s tags and SDKs, in order to provide data to Google and Facebook to use on the advertisers’ behalf to target ads and measure conversions of those ads.
- (a) Facebook aggregates data about users’ activities across all advertisers’ websites using Facebook Pixel and apps using Facebook’s SDKs to personalise ads.<sup>312</sup>
- (b) Whilst Google does not, in general, use advertisers’ analytics and Customer Match data for their own purposes or to help other advertisers, these data uploaded by advertisers are used to increase the performance of ads delivered on their behalf by Google. (For the minority of Analytics customers that have enabled the ‘share data with Google products and services’ setting, Google states that it can use this data to improve the Google Ads system tools.<sup>313</sup>) Furthermore, this does not apply to the data collected from publishers’ websites and apps that use AdSense and AdMob. Instead, user data collected from each publisher’s property are combined with data from other publishers’ properties that are also using these services, and collectively used to derive insights about users to personalise ads (subject to the user’s choice of privacy settings, including whether to activate personalised ads).<sup>314</sup>
294. For both Google and Facebook, these data are also combined with user data generated on Google and Facebook’s own properties, if users allow this (in their Google Account settings or in Facebook’s Off-Facebook Activity settings respectively). These combined data on users from multiple properties are used by Google and Facebook to inform targeting and ad selection.
295. As a result of Google and Facebook’s ability to target and deliver high performing ads, which is derived from their insight into users on their properties, third-party publishers are incentivised to use Google (and to a lesser extent Facebook’s) advertising services, allowing AdSense, AdMob, and Facebook Audience Network (FAN) ads to show on their properties because Google and Facebook can select ads (using the data and insights about users from their leading consumer properties) that perform better. Advertisers are willing to pay more for higher performing ads and so, in principle, more revenue is passed on to publishers.
296. For these reasons, many publishers of websites and apps include code (tags, pixels or SDKs) that allow Google and Facebook to track the behaviour

---

<sup>312</sup> Facebook Business Tool Terms, available [here](#).

<sup>313</sup> Google Analytics Help, Data Sharing Settings (available [here](#)).

<sup>314</sup> Google described this to us as a ‘Google data co-op’.

of their users to target ads and measure ad effectiveness. In using AdSense, AdMob and FAN, third-party publishers allow Google and Facebook to obtain even more data about consumer behaviour, including on non-Google and non-Facebook properties, which further reinforces their ability to target and deliver high performing ads.<sup>315</sup> In this way, tracking and market power are in feedback: publishers and advertisers depend on major players' services (which are backed by more user data and tracking), so let them track their users, giving them yet more presence and data.

### ***Estimates of the prevalence and prominence of tracking***

297. The amount of tracking currently occurring online can be measured in various ways, including measuring the quantity of third-party libraries (TPLs) inside websites and mobile apps, maintaining blacklists of known tracking domains, using graph analysis or analysing data sent over networks. In this section we summarise the results of a few different studies on the prevalence and prominence of tracking on websites and mobile apps.

298. The *prevalence* of a tracker can be defined as the number of websites and apps that it is present on. Prevalence does not account for the highly variable popularity of websites and apps, and therefore does not capture how many users are presumably impacted. Englehardt and Narayanan (2016) define a metric called *prominence*, which accounts for website popularity using the Alexa ranking.<sup>316</sup> However, not many studies adopt similar popularity-adjusted definitions when attempting to quantify the amount of tracking., We therefore caveat the reader about the difference between these two metrics while interpreting the results in this section.

### ***Websites***

299. Englehardt and Narayanan (2016) measured the presence of trackers in the Alexa top 1 million sites<sup>317</sup> in January 2016 and found that there is a long tail

---

<sup>315</sup> In addition, Google and Facebook also benefit from the standard indirect network effect that marketplaces (or entity that brings buyers and sellers together) experience. In addition to the benefits from better targeting and ad selection, publishers are also attracted to AdSense and FAN because many advertisers using Google and Facebook's advertising services means there is, all other things equal, more demand for inventory and a wider variety of creatives. Similarly, advertisers are also attracted to Google and Facebook's advertising services because many publishers using AdSense and FAN means there is more supply of impressions.

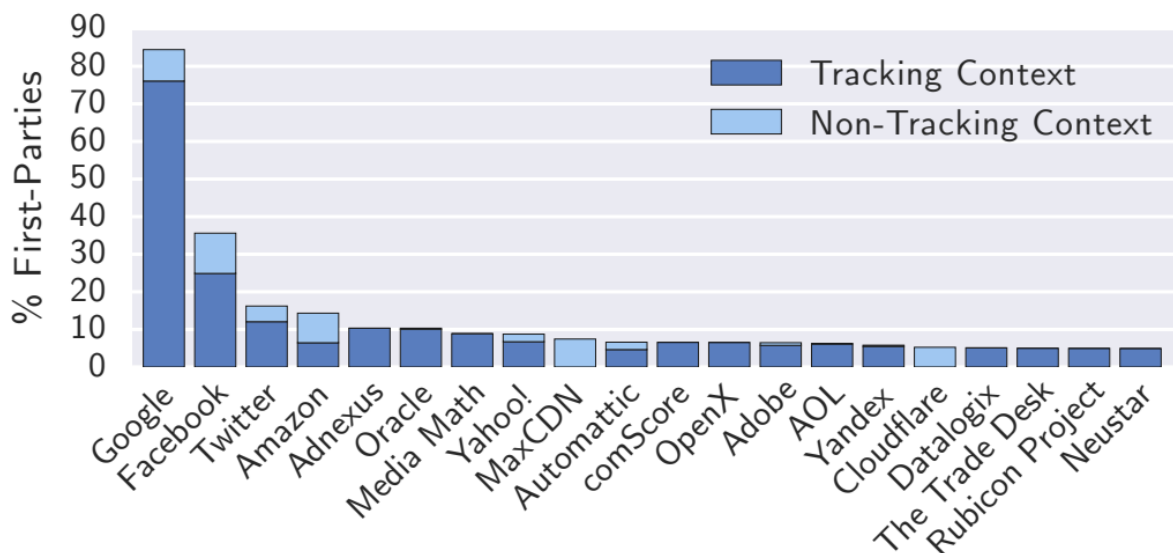
<sup>316</sup> They define the prominence of a third-party to be the sum of inverse popularity ranks for all websites the third-party is embedded in. The popularity rank is ordinal. The authors assume it is proportional to a website's audience, noting that actual user numbers would be more accurate but are unavailable. Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1388-1401). Available [here](#).

<sup>317</sup> [Alexa Top Sites](#). Since the Englehardt and Narayanan study in November 2016, Alexa stopped publishing the top 1m sites for free and provide no research access. Alexa now only provides the top 500 sites for free.



distribution in online trackers.<sup>318</sup> As shown in Figure G.11 below, they found Google, Facebook, Twitter, Amazon, AppNexus and Oracle are the only third-parties present on more than 10% of websites. Google was found to be in clear lead, with a third-party presence in approximately 85% of websites. The long tail distribution was confirmed in another study using a smaller dataset of the Alexa top 5,000 websites later in 2016.<sup>319</sup>

**Figure G.11: Organisations with the highest third-party presence on the top 1 million sites.**



Source: Englehardt and Narayanan (2016).

\*Note that the domains deemed in the 'tracking context' are considered to be those whose third-party resource would have been blocked by a consumer privacy tool.

† Note Adnexus is the domain of the company AppNexus.

300. Bashir and Wilson (2018) looked at adtech trackers involved in cookie syncing<sup>320</sup> The authors build an 'inclusion graph'<sup>321</sup> to investigate which advertising and analytics domains had the most third-party presence inside websites selling impressions.<sup>322</sup> The resulting graph is very dense and highly interconnected, with Google-owned domains occupying a central position that many other domains connect to, suggesting that large parts of the adtech industry depend on Google.<sup>323</sup>

<sup>318</sup> Englehardt, S., & Narayanan, A. (2016, October). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1388-1401). Available [here](#).

<sup>319</sup> Binns, R., Zhao, J., Kleek, M. V., & Shadbolt, N. (2018). Measuring third-party tracker power across web and mobile. *ACM Transactions on Internet Technology (TOIT)*, 18(4), 1-22. Available [here](#).

<sup>320</sup> Bashir, M. A. and Wilson, C. (2018). *Diffusion of user tracking data in the Online Advertising Ecosystem*. Proceedings on Privacy Enhancing Technologies ; 2018 (4):85-103

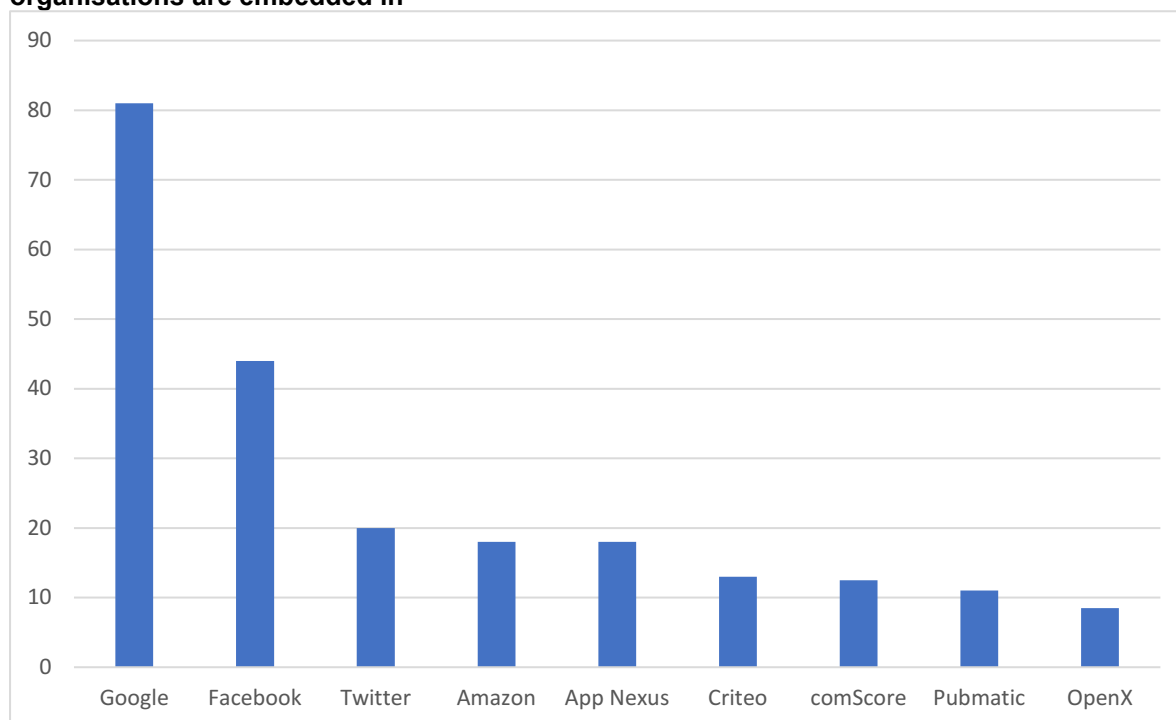
<sup>321</sup> This is a directed graph where an edge comprised of nodes  $A \rightarrow B$  indicate site A including B as a third party adtech tracker.

<sup>322</sup> Bashir constructed a dataset of 2 million impressions by crawling 150 popular e-commerce sites chosen manually from the Alexa top 1000 and several manually chosen product pages from 738 major e-commerce websites from the Alexa top sites in its Shopping category.

<sup>323</sup> Bashir and Wilson (2018) show many Google-owned domains in the top 10 most central tracking nodes. Specifically, Google domains have the highest betweenness centrality, that is it is the most common node along the shortest path between any two websites in the inclusion graph. The top 5 in order of highest betweenness

301. The overall picture of top trackers seems to not have changed much since the work by Englehardt and Narayanan (2016). Solomos et al. (2019) conducted a longitudinal study over the period September 2017 to April 2019, by crawling various subsets of the top 1 million Alexa websites to detect third-party domains<sup>324</sup>. Solomos et al. (2019) found that, over the period considered, tracker concentration increased – that is, fewer different trackers were embedded in more websites, and fewer HTTP requests directed to third-party domains. The authors note this latter finding may be due to publishers turning to well-known GDPR-compliant trackers. They found the top trackers domains retained their position over time, with Google firmly in the lead (in 81% sites) and Facebook (44%) second, as seen in Figure G.12. They note that the ‘almost immutable list of top trackers... points to the fact that the GDPR enforcement had no effect on them either in their importance in the web tracking ecosystem or their coverage across websites’.<sup>325</sup>

**Figure G.12: average (from 2017 – 2019) proportion of publishers that trackers from these organisations are embedded in**



Source: CMA, adapted from Table IV of Solomos et al. (2019)

centrality: Google-analytics.com, doubleclick.com, googleadservices.com, facebook.com, googletagmanager.com. Bashir found that adtech tracking domains were not balkanised into groups, but the inclusion graph was very dense and interdependent. Google’s centrality on this graph may suggest that many adtech players depend on them. Bashir, M. A. (2019). *On the Privacy Implications of Real Time Bidding* (page 79) available [here](#).

<sup>324</sup> These datasets include three before GDPR and three after GDPR. The full list of datasets used and their metadata are in Table 1 of page 2 in Solomos, K., Ilija, P., Ioannidis, S., & Kourtellis, N. (2019). Clash of the Trackers: Measuring the Evolution of the Online Tracking Ecosystem. *arXiv preprint arXiv:1907.12860*. Available [here](#).

<sup>325</sup> Solomos, K., Ilija, P., Ioannidis, S., & Kourtellis, N. (2019). Clash of the Trackers: Measuring the Evolution of the Online Tracking Ecosystem. *arXiv preprint arXiv:1907.12860*. Available [here](#).

## Mobile apps

302. A 2018 study by Binns et al. looked at third-party tracking in the Android ecosystem. The authors downloaded the source code of 959,426 popular apps<sup>326</sup> in the Play Store and cross-referenced the third-party libraries installed in them with known tracker blacklists<sup>327</sup>. They found that the largest players in the mobile tracking ecosystem do not differ substantially from those on the web. Figure G.13 shows Google is in lead, present in 88% of apps, followed by Facebook and Microsoft who are present in 42% of apps<sup>328</sup>. Several other companies – including Twitter (33%), Verizon (26%), and Amazon (17%) – have a significant presence. This suggests that the concentration in the mobile tracking market is not quite as stark as on the web, although the same players lead.

---

<sup>326</sup> Their sampling method was to use autocomplete of the [Play Store search](#) for all character strings up to length five.

<sup>327</sup> Binns, R., Lyngs, U., Kleek, M.V., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third Party Tracking in the Mobile Ecosystem. *Proceedings of the 10th ACM Conference on Web Science*. Available [here](#).

<sup>328</sup> Note that although Figure G.13 appears to have Microsoft as 22% prevalence, there is a mistake whereby LinkedIn should have been added to it as it has been a subsidiary of Microsoft since 2016.

Figure G.13: The most prevalent root parent tracking companies and subsidiaries.

<i>Root parent</i>	<i>% apps</i>	<i>Subsidiary</i>	<i>% apps</i>
Alphabet	88.44	Google	87.57
		Google APIs	67.51
		DoubleClick	60.85
		Google	39.42
		Analytics	
		Google Tag	33.88
		Manager	
		Adsense	30.12
		Firebase	19.20
		Admob	14.67
		YouTube	9.51
		Blogger	0.46
		Facebook	42.55
Liverail	1.03		
Lifestreet	<0.01		
Twitter	33.88	Twitter	30.94
		Crashlytics	5.10
		Mopub	2.51
Verizon	26.27	Yahoo	20.82
		Flurry	6.28
		Flickr	1.37
		Tumblr	1.22
		Millennialmedia	0.71
		Verizon	0.11
		AOL	0.06
		Intowow	<0.01
		One By AOL	<0.01
		Brightroll	<0.01
		Gravity	<0.01
		Insights	
		Microsoft	22.19
Bing	0.12		
LinkedIn	20.62	LinkedIn	20.62
Amazon	17.91	Amazon Web	11.57
		Services	
		Amazon	7.72
		Amazon	1.73
		Marketing	
		Services	
Unitytechnologies	5.78	Unitytechnologies	5.78
Chartboost	5.45	Chartboost	5.45
Applovin	3.95	Applovin	3.95
Cloudflare	3.85	Cloudflare	3.85
Opera	3.20	Adcolony	3.12
		Admarvel	0.09

Source: Binns et al. (2018) *Third Party Tracking in the Mobile Ecosystem*. Available [here](#).

\* Note that LinkedIn is a subsidiary of Microsoft (since 2016) and Crashlytics is a subsidiary of Google not Twitter (since 2017).

303. Another study focused on mobile tracking is Razaghpanah et al. (2018). The authors crowdsourced data from their app Lumen, including traffic network metadata characterising data flows from 14,599 apps<sup>329</sup> installed on 11,384 Lumen users' devices.<sup>330</sup> This study found that 16 of the 20 most popular advertising and tracking third-party services were connecting to Alphabet-owned domains. They found that Alphabet was present in over 73% of apps, Facebook in 31% and Verizon in 13% of apps. We note that these results are similar to those found by Binns et al. (2018) above.
304. The proprietary nature of Apple's iOS makes investigation and research of trackers more complicated. We were thus unable to assess the prevalence of tracking on iOS Apple devices. The 'walled-garden' architecture of iOS arguably makes the ecosystem harder for trackers to embed into in the first place (see for example the earlier discussion on pre-installed apps).

### *Mobile and web comparison*

305. Razaghpanah et al. (2018), cited above, also included a comparison of web and mobile tracker prevalence by domain.<sup>331</sup> On the web, google.com is present on around 70% of websites, and google-analytics.com on around 60% of sites. The most prominent mobile trackers are also Google-owned domains. In addition to google-analytics.com, prominent Google advertising domains include googlesyndication.com<sup>332</sup> and doubleclick.net.
306. It is worth noting that 16 of the 20 most prevalent advertising and tracking domains are Google owned. The Alexa top 1,000 list was used for websites here, which accounts for prominence somewhat, but the mobile apps were just prevalence counts – although we note their data source was from a large representative population, so popularity may have been endogenous and therefore somewhat accounted for.

---

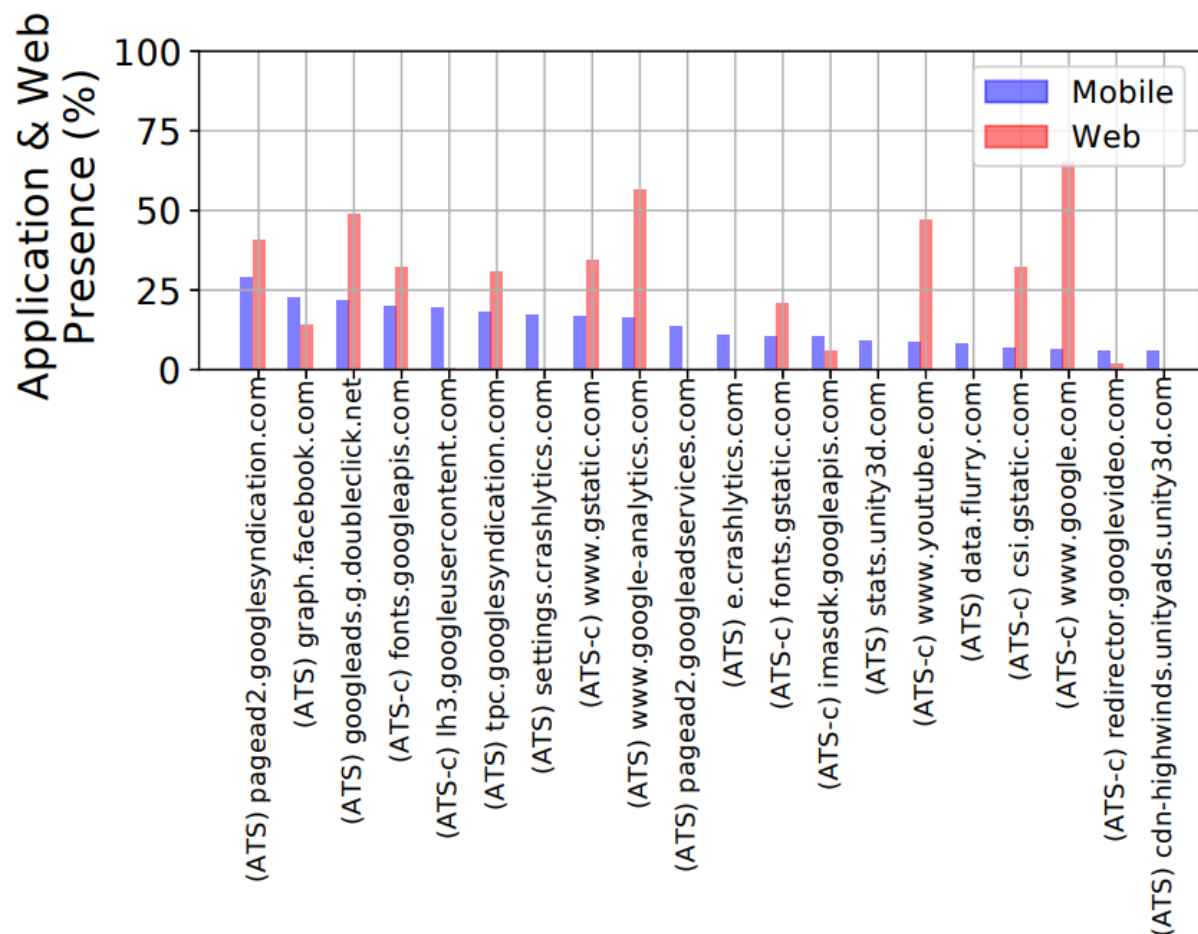
<sup>329</sup> Razaghpanah et al. consider these apps to be representative of those used by the average mobile users as 48% of them have more than 1 million installs and 71% of them are listed in Google Play's Top-50 charts for the US, Spain, Germany, India and the UK. Lumen is able to capture network traffic data and associate it to processes running (apps, in sandboxes) on Android due to their app which the user consents to give privileged access which inserts itself as middleware between apps and the network interface, leveraging the Android VPN permission.

<sup>330</sup> Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., & Gill, C. K. P. (2018). Apps, trackers, privacy, and regulators. In *25th Annual Network and Distributed System Security Symposium, NDSS* (Vol. 2018). Available [here](#).

<sup>331</sup> Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., & Gill, C. K. P. (2018). Apps, trackers, privacy, and regulators. In *25th Annual Network and Distributed System Security Symposium, NDSS* (Vol. 2018). Available [here](#).

<sup>332</sup> Google told us that GoogleSyndication.com is a domain, owned by Google, used for storing and loading resources including ad content. It records user interaction with AdSense ads (such as which ads the user clicked on, and what category the ads fall under). Based on this information, an inferred user profile is created including the user's interests, possible intended purchase behaviour and other characteristics.

Figure G.14: Presence of the most popular advertising and tracking services in the apps on Lumen users' devices and the Alexa Top 1,000.



Source: Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., & Gill, C. K. P. (2018). Apps, trackers, privacy, and regulators. In 25th Annual Network and Distributed System Security Symposium, NDSS (Vol. 2018). Available [here](#).

307. The Razaghpanah et al. study also found cross-device tracking in 39% of the tracking SDKs that Razaghpanah et al. (2018) found in their sample of Android devices were also present on at least one Alexa top 1K website, which is suggestive of the potential for these SDKs to engage in cross-device tracking. They also found 17 of the top 20 trackers having both a large web and mobile presence (with Google in first place and Facebook in second). When considering prevalence and prominence of trackers inside websites and apps for the purposes of assessing how tracking and market power interact or explaining the extent of data processed on users, it is useful to include cross-device prevalence because a very accurate picture of a user's life can be built with cross-device information.

308. In sum, tracker measurement studies from 2016 to 2019 characterise the scale, distribution and major trackers in the web and mobile tracking markets for advertising and analytics show that Google is firmly in lead, with Facebook

in second place. Google’s strength is more pronounced on web than on mobile.

### *Google’s coverage of UK population for certain identifiers*

309. Given the prominence of Google’s trackers in the web and mobile ecosystems, the CMA asked Google to provide further information on their coverage. In Table G.4 we give those numbers that we consider to be representative of the UK consumer population.
310. We asked Google to provide the number of UK users for whom they held various identifiers on.

**Table G.4: Indication of the number of UK users whom Google collects identifiers on**

<i>Identifier</i>	<i>Number of UK users</i>
Mobile advertising ID (MAID)*	[40,000,000 – 50,000,000]
IP address†	[60,000,000 – 70,000,000]
Location‡	‘Half to two-thirds of Android users’
IMEI	‘Most Android users’
MAC address	‘Most Android users’

Source: Google

\* Daily average of distinct Advertising IDs (AdIDs and IDFAs) in the UK (based on user IP address) recorded by GA ‘Gold’ app logs (which cover GA for Firebase and ‘App + Web’ properties) from 13-17th March 2020.

The daily average of distinct Advertising IDs in the UK recorded in Google Analytics Classic app logs (which covers Analytics and Analytics 360) covering the period 11-17th March 2020 was [10,000,000 – 11,000,000].<sup>333</sup>

† Google Search users. Google told us that it will likely receive the IP addresses of the vast majority of Google Search users.

‡ This is the approximate proportion of Android users who have opted into Location Services.<sup>334</sup>

311. We can see that Google hold the IP address and the MAID (a unique mobile device identifier) for most of the UK population. Furthermore, we note that half to two-thirds of Android users opt into Google’s Location Services.
312. For context, Google set out that its total active UK logged-in user base over the 28-day period 21 July 2019 to 19 August 2019 was [50,000,000 – 60,000,000]. This is relevant insofar as any of the identifiers above that Google collect can be matched to Google’s internal account IDs when a user is logged in, although Google told us that their ‘advertising systems do not collect or make use of IMEI, MAC address or other identifiers, excluding narrow fraud or abuse situations’. Additionally, Google told us that in 2019

<sup>333</sup> We expect these figures to overlap somewhat, but the true number is larger than any one of them.

<sup>334</sup> Google Location Services is an Android feature. It should not be confused with Location History, which is collected from Google Account users, on both web and mobile (for both Android and non-Android users). More is available in Google’s FAQs on Location Services available [here](#).



[10-20%] of apps accessed by users in the UK on Android use the Google Sign-In functionality.<sup>335</sup>

## **Recent and near-future developments in tracking, web standards, and privacy-enhancing technologies**

313. Digital advertising platforms track consumers and use data about consumers to accomplish two main goals, which are discussed above and in more detail in Appendices F, M, O:
- (a) To target ads to consumers, so that consumers are shown ads that they are likely to be interested in (behavioural targeting); and
  - (b) To ensure that ads are viewed by the right number and kinds of people, avoiding fraud and invalid traffic (verification), and to evaluate the effectiveness of campaigns, by relating ad exposure to conversions, ideally across different devices (attribution and evaluation).
314. These activities (behavioural targeting, verification, and attribution) contribute to the efficiency of the digital advertising market. Furthermore, as discussed in Appendix F, if competitors had access to the data needed for these activities, they may be better able to compete with large platforms in providing services to users and for advertisers.
315. However, currently each of these activities involves the gathering, remote processing, and (sometimes) transfer of large quantities of user data across publishers, advertisers, platforms and intermediaries in the digital advertising supply chain. This data is often personal data within the meaning of GDPR, which raises data protection and privacy issues.<sup>336</sup> In particular, the data (including personal data) is sent to a potentially very large number of third parties, which might not align with users' knowledge and expectations about how their data is shared and used. These concerns have led to measures and proposals that, whilst potentially enhancing privacy, could also potentially reduce the efficiency of digital advertising and harmed the ability of independent, non-vertically integrated adtech providers to compete with large incumbent platforms like Google and Facebook.
316. A crucial question is whether and to what extent these activities that contribute to the efficiency of the digital advertising market can be performed in a way which better protects privacy and better facilitates competition by

---

<sup>335</sup>There were 300,000 - 400,000 apps accessed by UK users in 2019 on Android, 60,000 - 70,000 of which used Google sign-in functionality according to Google.

<sup>336</sup> These concerns discussed in Chapter 4 and are set out in the ICO's [Update report into adtech and real time bidding](#), 20 June 2019.

preserving ability of smaller firms to operate effectively in the relevant markets.

317. This section:
- (a) discusses the potential impact of ongoing and upcoming restrictions of a significant technology for tracking, the third-party cookie; and
  - (b) outlines several proposals in the web standards community and future developments in privacy-enhancing technologies (PETS), which would reduce the extent of tracking and data collection for digital advertising by shifting a significant proportion of the data processing to the device itself, which could help better protect privacy whilst preserving some of the efficiencies from current digital advertising.

### ***Control over web standards and relative dependence on third-party cookies***

318. Google is at the centre of a complex ecosystem (set out in Appendix E) including the browser with the highest share of use in the UK (Chrome) and the open source projects that provide code underlying most browsers (Chromium) and mobile operating systems (Android). Google has an outsized impact on the standards for how technology develops, in turn determining who can collect data. In this way Google may favour architectures that maintain its position in being a gatekeeper of user data, with other companies becoming dependent on and vulnerable to decisions it makes. Its influence over web standards via the Chrome browser has a direct impact on the adtech ecosystem, as its proposed changes to Chrome have the potential to directly prevent rival adtech companies from implementing their own targeting and measurement solutions.
319. In January 2020, Google announced its intention to phase out support for third-party cookies on Chrome within two years.<sup>337</sup> (As discussed above, the web standards community generally defines a cookie as first-party when the registrable domain<sup>338</sup> of the page visited by the user matches the registrable domain of the cookie. If the registrable domains do not match, then the cookie is considered third-party to the page.)
320. As explained in previous sections, third-party cookies currently play a very important role in the adtech ecosystem. As the current principal means of achieving common identification of users, albeit imperfectly, third-party cookies are a fundamental building block of open display advertising and

---

<sup>337</sup> Chromium Blog, 'Building a more private web: A path towards making third party cookies obsolete', available [here](#).

<sup>338</sup> Registrable domain is effective top-level domain plus one additional label (eTLD+1). For instance, 'www.google.com' and 'news.google.com' share the same registrable domain.

make possible the flow of data about users through the digital advertising ecosystem needed to target advertising and measure conversions.

321. It is important to place Google's proposal to deprecate third-party cookies within the wider context of an interrelated set of proposals by Google called Privacy Sandbox.<sup>339</sup> Google is adopting a phased approach to limit the use of third-party cookie over two years, whilst actively developing and testing alternatives to replace the functionality served by cross-site tracking and third-party cookies. These functions include ad targeting (including interest-based targeting and remarketing) and ad conversion measurement, but also combating spam and fraud, and federated log-in. The Privacy Sandbox and the specific proposals within it are discussed in detail in the sections below.
322. Google told us that Chrome's deprecation of third-party cookies is conditional. If, by 2022, Google judges the other proposals in Privacy Sandbox to be overall unsuccessful or insufficiently developed, it will modify its approach to (and may delay or suspend) the deprecation of third-party cookies on Chrome.
323. Google's announcement is significant because Chrome is the most popular browser in the UK,<sup>340</sup> but it has been widely anticipated. Safari and Firefox have already implemented similar measures unilaterally to prevent cross-site tracking of their users.
324. Several adtech providers told us that they were significantly impacted by Apple's decision to implement Intelligent Tracking Prevention (ITP) on Safari in September 2018, which limited the ability of adtech providers to implement third-party cookies on Apple's Safari browser and therefore to access and collect data on Safari users. Our main evidence for the short-term impact of the removal of third-party cookies is our analysis of the data for Google's cookie RCT, discussed in Appendix F.
325. In principle, Google's open display activities will also be affected by the deprecation of third-party cookies. Google stated that once third-party cookies are no longer supported, it will generally not be able to associate ad requests from third-party sites with Google Account-level data for individual users, or to use the data received in such ad requests to recognise users' cross-site activity. This means that, if Privacy Sandbox were to launch as currently

---

<sup>339</sup> For example, as part of Privacy Sandbox, Google is also developing a proposal for First-Party Sets (see the [First-Party Sets GitHub page](#)), which will allow related domain names owned by the same entity (eg apple.com and icloud.com) to declare themselves as the same first party, so that the deprecation of third-party cookies does not sever the ability of commonly owned first-party domains to set cookies on each other. There is an open question about whether this is desirable – [Apple](#) and [Mozilla](#) have raised several concerns about this proposal, such as whether users will be aware of these affiliations between domains, and incentives for publishers to form and personalise first-party sets.

<sup>340</sup> According to StatCounter, it had a market share of approximately 50% in October 2019. Available [here](#). For a discussion of the methodology used by StatCounter, see Appendix C.

designed (see section below on ‘Privacy-enhancing technologies and proposals’):

- (a) Like others in the industry, Google will generally be unable to use data generated by users’ activities on non-Google websites to personalise advertising to those same users on different non-Google websites. However, this does not preclude advertisers and publishers from uploading data they collect (using first-party cookies and volunteered by users) to Google, using Google’s Customer Match service (explained in more detail above, in the section on Remarketing lists and Customer Match). Google could match this data using other identifiers (such as names, email addresses, mailing addresses and phone numbers collected directly from users) and use the combined data to target ads on Google’s own properties.
- (b) Taking the stated intention of the deprecation of third-party cookies and Privacy Sandbox at face value, which is to limit cross-site tracking, Google will generally not be able to use the insights that it obtains from users on its properties, which it associates with its first-party cookies or Google Accounts, to personalise ads for users viewing impressions on non-Google properties. However, we note that it is possible to circumvent blocks on third-party cookies, by asking advertisers and publishers to implement equivalent tracking code using first-party cookies.
  - (i) For instance, Google Analytics tags are currently implemented using first-party cookies.<sup>341</sup> (See section above on Google Analytics, Floodlight, and Google Tag Manager.)
  - (ii) To take another example, Facebook Pixel collects data from non-Facebook properties which is used for Facebook’s advertising services, and websites can implement Facebook Pixel using first-party cookies.<sup>342</sup> This means Facebook Pixel can work with browsers blocking third-party cookies.<sup>343</sup> (See section above on Facebook Pixel.)

326. We make a number of further observations about the likely impact of the deprecation of third-party cookies and Privacy Sandbox, if it is successfully implemented and workarounds that reintroduce cross-site tracking are prevented:

- (a) First, targeting using publishers’ first-party data and authenticated user data does not require cross-site tracking and is unaffected by the demise

---

<sup>341</sup> See [Google Analytics Cookie Usage on Websites](#).

<sup>342</sup> Facebook, [About cookie settings for Facebook pixel](#).

<sup>343</sup> See, for instance, Clearcode, [What Facebook’s First-Party Cookie Means for AdTech](#), available [here](#).

of third-party cookies. Therefore, large incumbent platforms with leading consumer-facing services like Google and Facebook are significantly less dependent on third-party cookies for delivery of high-performing targeted ads and continued advertising revenues than, for instance, small publishers with free-to-read content that does not require log-in.<sup>344</sup> Multiple stakeholders, in response to our interim report, stated that the removal of third-party cookies in particular would further entrench of Google's (and potentially Facebook's) adtech solutions, which have access to large quantities of first-party data. This was raised by research stakeholders (Prof. Geradin and Katsifis) as well as publishers (DMG, NMA). Oracle pointed out that some intrusive data collection practices by large platforms, which are not dependent on third-party cookies and cross-organisation tracking, would not be hindered in this scenario.

- (b) Second, contextual advertising also does not require cross-site tracking. It is widely anticipated that advertisers will return to spending larger proportions of their budget on contextual advertising. Adtech providers are already researching and developing new technologies to enhance the effectiveness of contextual advertising, for instance by deploying natural language processing and computer vision to further improve their ability to automate rapid interpretation of the content of web pages and apps at scale. This has the potential to enable contextual targeting that is based on more nuanced understanding of the language, images and the sentiment of the context, which could help advertisers to avoid the brand safety issues of more naïve approaches for contextual targeting such as keyword matching.
- (c) Finally, if successfully implemented, Google's main Privacy Sandbox proposals, Federated Learning of Cohorts (FLoC)<sup>345</sup> and TURTLEDOVE,<sup>346</sup> may still permit some third-party personalised advertising (interest-based advertising and remarketing), albeit at a greater level of coarseness of targeting and measurement. However, those proposals will also turn Chrome (and Chromium browsers in general, if other browser like Edge and Firefox decide to implement the relevant code into their own browsers) into the key gateway for adtech. These proposals are discussed in more detail in the section below on Privacy Enhancing Technologies. It is likely, therefore, that Google's

---

<sup>344</sup> For example, Facebook stated that a key design advantage of Facebook's Conversion Lift tool is its "single-user login" feature which tracks users via a single-user login across devices and sessions, which is a significant improvement over more common cookie-based approaches.

<sup>345</sup> FLoC allows the user's browser to withhold personal data about cross-site browsing history used to infer interests but instead to reveal k-anonymous data about the interests of the user's cohort of k users.

<sup>346</sup> TURTLEDOVE will mean that browsers accept requests from advertisers to show users remarketing ads ('bids') based on signals of potential interest by the user and, when an opportunity to show an ad later arises, the browser will run an 'auction' with the pre-submitted bids and determine which retargeting ad to show the user.

position in the adtech ecosystem will remain central. Market participants may be concerned that, under these proposals, Chrome would have the ability to use its position to favour Google's own adtech intermediation services.

### ***The potential benefits of privacy-enhancing technologies (PETs)***

327. In the current system, data generated by users can be used to track their identities across online and offline activities, serve individually targeted ads, and measure how these ads affect their behaviour. For these purposes, data gathered from users' devices is processed remotely by various actors in the supply chain.
328. Privacy-enhancing technologies (PETs) are a class of technologies that seek to mitigate privacy risks associated with the collection, transfer, and analysis of data, while still allowing for useful results to be obtained from the data. PETs encompass a wide range of approaches, with different degrees of maturity and applicability.
329. A particular type of PETs is client-side PETs. Approaches of this type aim to shift a significant proportion of data processing to the client side (for example, the user's device itself), reducing the amount and granularity of the information that gets transferred away from it. In this way, the ability of adtech providers to identify and profile individual users during their online activity is potentially curtailed.
330. The remainder of this section focuses on approaches based on client-side PETs. This is because most existing PET proposals in the digital advertising ecosystem are concerned with on-device processing.
331. Client-side PETs preserve some of the ability for advertisers to provide ads that are targeted to users' interests. The fundamental difference is that a higher proportion of the processing (eg assigning users to segments or matching impressions to ads) happens on the user's device, rather than remotely.
332. Verification, measurement, and attribution are also potentially achievable in a privacy-enhancing manner, by also shifting the matching between exposure and conversion events to the device, and only sending anonymous and or aggregate attribution data to advertisers, rather than relying on individual-level tracking.
333. These approaches can thus potentially be implemented without compromising the free ad-supported model that underlies a significant proportion of online content creation by publishers.

334. Furthermore, privacy-enhancing approaches could reduce or eliminate the incentives leading to large scale data collection, storage, and resale by Data Management Platforms (DMPs), which can constitute a significant challenge to privacy.
335. Finally, they would not rely on unique identifiers such as the Mobile Advertising ID, which can facilitate tracking of users by third parties.
336. Whilst the client-side privacy-enhancing technologies and approaches we discuss here may result in significant gains to privacy without sacrificing too much efficiency, based on our current understanding and the evidence reviewed so far, they do not completely remove or fully overcome the trade-offs between privacy, efficiency, and competition that seem to be inherent in digital advertising. As discussed in Appendix T, any remedy intervention must consider these three dimensions jointly.

### ***Privacy-enhancing technologies and proposals***

337. There is currently an active debate within the web standards community about restricting cross-site tracking, and mitigating the impact that this will have on current digital advertising use cases that rely on it. There are many proposals,<sup>347</sup> and we cannot adequately cover all of them in this section. The purpose of this section is to set out a few of the more prominent proposals and ideas, which informed our recommendations and should inform the design of future remedies and regulations at the interface of competition, privacy, and the use of data for digital advertising and online platforms in general.

#### *Privacy budgets*

338. At the time of writing, there is an active discussion within the web standards community about curtailing browsers' vulnerability to fingerprinting by limiting the amount of information that browsers expose to websites, whilst balancing the need for websites to get access to information in order to provide useful functions, within the framework of a 'privacy budget'.<sup>348</sup> The general idea is for browsers to measure how much identifying information (or entropy) is given away when it exposes any given piece of information to a website. This measure can then be used by browsers to constrain websites to a predefined entropy budget, so that they are incentivised to only ask for what they need.

---

<sup>347</sup> The W3C Web Advertising Business Group has a fairly comprehensive list of advertising use cases that currently depend on cross-site data sharing, and an assessment of whether these use-cases may be supported by Chrome Privacy Sandbox proposals and Safari, as well as community proposals (available [here](#)).

<sup>348</sup> Google, 'Combating Fingerprinting with a Privacy Budget'.



Websites that exceed these limits can be prevented from accessing more data.

339. In our view, the efficacy of privacy budgets will depend on behavioural factors – in particular: (i) the extent to which browsers enforce budgets by default, and (ii) the design of any user interfaces and information about privacy budgets. The goals of privacy budgets may be undermined if websites can routinely prompt users to grant permission to go ‘over budget’. As discussed in the section above on ‘Transparency and Consent Framework (TCF)’, users rarely read privacy information before using a website or app. It is possible that many typical users will consent, without due consideration, to lifting any privacy budget in order to quickly access the website or service, if they are given the opportunity to do so.<sup>349</sup>

*Client-side privacy-enhancing technologies (on-device processing and edge computing)*

340. The main difference that sets client-side privacy-enhancing approaches apart from the current models is the increased focus on processing data *on-device*. Raw information about the user and their online interactions, which might include personal data and special category data, is only accessed and processed by the device itself, instead of being transmitted in its raw form to be processed elsewhere.
341. Major tech companies are currently offering software developers the capability to access advanced computing resources on their devices, including CPUs, GPUs, and AI-specific hardware components. Developers can build machine learning models and deploy them within their apps so they run on the device itself, with full or partial access to the device’s capabilities.<sup>350</sup>
342. On desktops, services are generally accessed via the browser. Thus, privacy-enhancing technologies would likely be implemented as part of browser software. On mobile, many services are accessed by apps outside of browsers – which would require a more device-wide approach.
343. While raw user data might not leave the device, there are instances in which it might be desirable to make other types of user-generated data available as a user interacts with online services. In such cases, a valid privacy-enhancing approach must still make it impossible for other actors communicating with the browser/device to identify the individual behind these interactions. To this purpose, additional privacy requirements can be imposed – such as k-

---

<sup>349</sup> For instance, potentially in the form of [User Agent Client Hints \(UA-CH\)](#).

<sup>350</sup> See for example Apple’s [Core ML](#) framework, or Google’s [Coral](#).

anonymity (for individual data being broadcast by the browser)<sup>351</sup> or differential privacy (for statistics or models created using individual data).<sup>352</sup>

344. If on-device processing were feasible and became a standard default (either through effective competition on user privacy between device manufacturers, or through mandatory regulations), one of the notable advantages would be to place less burden on consumers. By not requiring users to actively affirm consent on a continuous basis and reducing their need to familiarise themselves with ways to preserve their privacy online, it might reduce consent fatigue.

### *Privacy-enhancing behavioural targeting*

345. Behavioural targeting aims to serve ads to specific users based on their observed and inferred characteristics and interests. Typically, behavioural targeting exploits the availability of large quantities of individual-level data on characteristics (eg demographics, browsing history, search terms).
346. The effectiveness of behaviourally targeted personalised ads is assessed by associating users with data about ad exposures and conversion events (eg clicks on the ad, purchases and subscriptions). Privacy-enhancing approaches to verification, attribution and evaluation of ads are discussed in a later section.
347. In general, considering the positions of major platforms on potential privacy-preserving behavioural targeting:
- (a) Safari does not support and appears to have no intention of supporting personalised advertising use cases that involve the use of information and signals generated in other contexts (ie other than the current webpage or app that the user is interacting with) to select ads, such as users' behaviour on webpages and apps that they have previously used. For instance, Safari provides no support for retargeting, lookalike targeting, or frequency capping.
  - (b) By contrast, Google (and Facebook via the web standards community) are very active in developing proposals for Chromium where these use cases are supported somehow in a privacy preserving way.

---

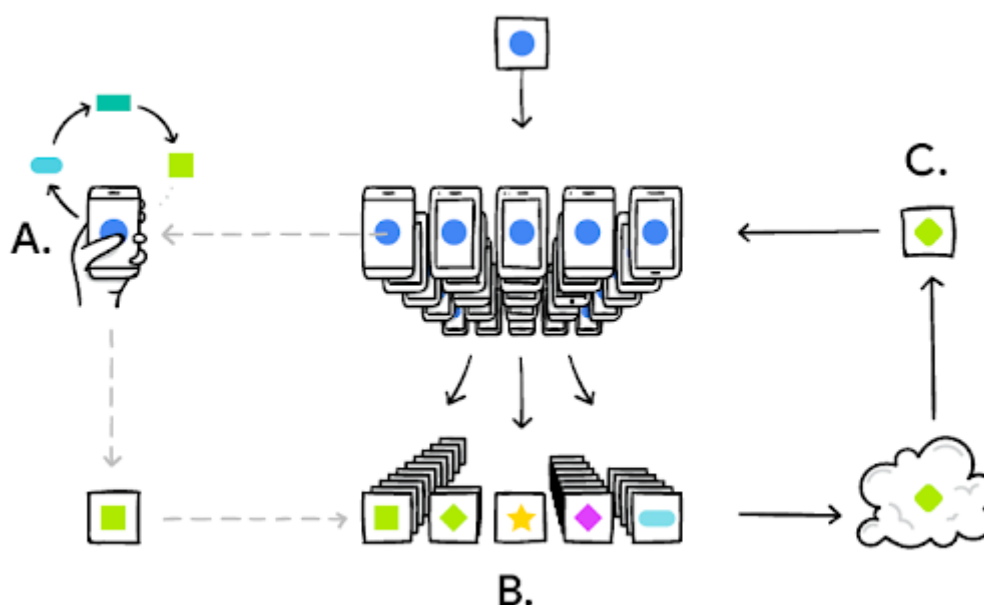
<sup>351</sup> K-anonymity is a framework that aims to achieve K-anonymity of individual data by ensuring that an individual's data is indistinguishable from at least  $(k - 1)$  others' (see L. Sweeney (2002), [k-Anonymity: A model for protecting privacy](#). International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570).

<sup>352</sup> Differential privacy is a security concept 'which means that, when a statistic is released, it should not give much more information about a particular individual than if that individual had not been included in the dataset' (Royal Society (2019), [Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis](#), p.13).

## Federated Learning

348. Before discussing various proposals for ‘private behavioural targeting’, this section provides some key ideas on federated learning, which will be necessary to understand one of the proposals in Google’s Privacy Sandbox.
349. Advertisers use machine learning (ML) approaches to train models that predict the likelihood of conversion events based on observed characteristics. These models can in turn be used to predict the likelihood of conversion for a previously unseen user with similar characteristics. Users with higher conversion likelihood in a specific context will be assigned a higher value, and the advertiser will be willing to pay more to show them ads.
350. Typically, ML models for targeting are developed, trained, and refined in a centralised manner; data is gathered from users, processed on remote servers, and then the results are used to decide which ads to serve (or how much to bid for the chance to serve an ad) to a newly observed user interactions.
351. A possible alternative approach to centralised model training is federated learning (FL). The main intuition behind FL is that the training of ML models can occur in a decentralised manner across multiple devices, instead of a single centralised instance.

Figure G.15: Example federated learning flow



Source: Google.

352. Consider the example in Figure G.15. The current ML model (blue circle) is sent to a user’s device. The device then uses the data generated locally by the user’s behaviours and interactions to improve the model (A) and produces

a partial update to the current model. Updates from multiple users are encrypted (B) and securely transmitted to the cloud, where they get decrypted and aggregated into a new model. Throughout the process, the user data on which model training is performed never leaves the device. Furthermore, the model updates that do leave the device are encrypted and anonymised so that they cannot be associated with any individual user.

353. Currently, Google implements a FL approach in multiple ongoing applications – from improving its predictive keyboard, to enhancing mobile vision, to automatic captioning of video content.

### *Chromium Privacy Sandbox – FLoC and TURTLEDOVE*

354. As part of Chromium’s ‘Privacy Sandbox’, Google has put forward a proposal known as Federated Learning of Cohorts (FLoC) aimed at reducing the privacy footprint of behavioural targeting with the use of FL.<sup>353</sup> This proposed approach, still at an early stage, would operate through any browser that chooses to implement this. The browser would use a federally trained on-device model to assign users to segments (‘flocks’) with similar browsing habits, and the user’s ‘flock name’ can then be sent by the browser as an HTTP header to websites and passed on to adtech providers for behavioural targeting. If these clusters are large enough, the developers claim that privacy by k-anonymity would be ensured.<sup>354</sup>
355. This type of approach uses an on-device model to assign users to segments, and the data used to train this model stays on the device. However, data about users’ membership to segments does leave the device and is accessible to websites. While less disclosive than cookies per se, segments still contain potentially personal information about individual interests, including sensitive categories. Furthermore, repeated queries to the browser to access a user’s segments can be used for tracking or fingerprinting purposes, in conjunction with other information such as IP addresses. To be truly private, this type of approach would have to be coupled with another layer of privacy-enhancing technology.
356. TURTLEDOVE<sup>355</sup> is a proposal for advertisers to show ads to ‘potentially interested’ users who have previously interacted with the advertiser or ad network. This proposal is aimed at replicating the functionality of user-lists and

---

<sup>353</sup> See the [explainer for Federated Learning of Cohorts \(FLoC\)](#).

<sup>354</sup> FLoC may be able to provide limited support for lookalike targeting or audience expansion, since each flock would be a natural audience, and it may be possible to analyse the aggregate behaviour of different flocks to find similar or adjacent flocks in some meaningful way that has value to advertisers and publishers. Facebook also has a proposal on [Privacy Preserving Lookalike Audience Targeting](#).

<sup>355</sup> In keeping with the avian theme of the acronyms of other proposals in Privacy Sandbox, TURTLEDOVE stands for ‘Two Uncorrelated Requests, Then Locally-Executed Decision on Victory’.

custom audiences, for behavioural, interest-based marketing and retargeting (discussed above and in Appendix F).<sup>356</sup> The key ideas are that:

- (a) After a user first visits an advertiser's website, the advertiser ask the browser to join one or more interest groups for a set amount of time. The key difference with current approaches is that browsers keep the information about which interest groups the user is member of, rather than the advertiser. Interest groups may be defined by advertisers. For instance, an advertiser may define an interest group comprised of all users that visited a product page featuring a pair of running shoes. To prevent micro-targeting, browsers may need to prevent interest groups that are too small.
- (b) Later, the browser requests ads targeted at the interest groups the user is part of. These ads may be cached for later use. This '**interest-group ad request**' does not contain any information about the user or what web page the user is visiting. It is likely to also be made in advance of the impression arising, so the response must include some bidding logic (including frequency caps, and logic that can process signals about the context including ad slot size/formats and brand safety) to generate a bid in response to any given context for an impression that arises in future. Crucially, the logic must be executed purely locally on the browser, without any access to any external system.<sup>357</sup>
- (c) When the user is visiting another website and an impression arises, the browser also requests ads which may be targeted using information about the context and the publisher's first-party data about the user in the request, similar to how bid requests are sent out now but, crucially, without any user IDs or user groups (a '**contextual ad request**').
- (d) The 'interest-group ad request' and the 'contextual ad request' must be kept independent and uncorrelated, so that they cannot be linked to the same person. Therefore, they have to be sent at different times (hence why interest-group ad requests must be made in advance), in addition to a host of other supporting conditions (such as effective prevention of

---

<sup>356</sup> PETREL (Privacy Exclusion Targeting Rendered Exclusively Locally), an alternative proposal from Facebook to TURTLEDOVE, adapts the key ideas in TURTLEDOVE to accommodate exclusion targeting (ie negative interest groups, for which advertisers will not show ads to, such as those users that have already made a recent purchase). PETREL is discussed [here](#).

<sup>357</sup> This will disrupt advertisers' ability to have fine-grained, real-time control over their campaign pacing and budgets, as advertisers would not be able to update bidding logic already stored in browsers, until browsers decide to request an update. An alternative proposal by Criteo called SPARROW (Secure Private Advertising Remotely Run on Webserver) could address this problem, as interest-group creatives and bidding logic are stored with a handful of known, centralised Gatekeepers instead of in theoretically unknown users' browsers. (For more details on SPARROW, it is discussed [here](#).)

fingerprinting, and limits on the use of multiple interest groups in a single interest-group ad request).

- (e) The browser then conducts an auction on-device to decide which ad (if any) to show to the user, using information from both the interest-group bid responses and contextual bid responses.
- (f) The winning ad is rendered in the browser, in a way which does not leak information to the surrounding webpage (such as in an opaque iFrame). This could raise challenges for verification (reporting) and attribution, which are addressed in the next section.

357. As mentioned in the previous section on ‘Control over web standards and relative dependence on third-party cookies’, both the FLoC and the TURTLEDOVE proposals place the browser in a vital gatekeeper position for the adtech ecosystem. We note that Criteo has published a competing proposal called SPARROW, which maintains the same privacy-enhancing objectives as TURTLEDOVE, but several key roles would be performed by a completely independent ‘Gatekeeper’) instead of the browser. This Gatekeeper cannot have any other role in the adtech ecosystem. The Gatekeeper holds the interest-group bid responses (the creatives and bidding logic) that advertisers submit; it runs interest-group auctions; and it handles the rendering of ads, but it does not receive or hold user-level information. There can be multiple competing Gatekeepers, and their independence could be enforced by audit and regulation either by an industry consortium or regulators.<sup>358</sup>

### *Privacy-enhancing verification, measurement and attribution*

358. Users’ browsing data plays a critical role in verification, measurement, and attribution tasks for digital advertising. In addition, advertisers assess ad exposure and link it to conversion events to measure the effectiveness of campaigns, using various techniques to reconstruct consumer journeys across websites and devices.

359. Some of these techniques result in privacy-invasive accumulation and transfer of users’ personal data, and have been disrupted by Apple (with WebKit’s Intelligent Tracking Prevention)<sup>359</sup> and Mozilla (with Firefox’s Enhanced

---

<sup>358</sup> For more information on SPARROW (Secure Private Advertising Remotely Run on Webserver), please see [here](#).

<sup>359</sup> See the [collection of privacy blogs on WebKit](#), which include the history of and latest updates to Intelligent Tracking Prevention.

Tracking Protection)<sup>360</sup> equipping their browsers with default options to curtail common web tracking approaches, such as tracking cookies.<sup>361</sup>

360. This section sets out some of the current proposals on privacy preserving ad click measurement, attribution and reporting.

#### *WebKit Private Click Measurement*

361. Through its WebKit browser engine, Apple has recently put forward a new on-device technology proposal aimed at allowing attribution of ad clicks without the need to track individual users, called Private Click Measurement.<sup>362</sup> This approach stores information on ad clicks, intended ad click destinations, and conversions on the user's browser. Ad campaigns and conversion events are denoted by the publisher (the ad click source) using 'small' identifiers (up to 6 bits each, or 12 bits in total),<sup>363</sup> which contain too little information to be used as cross-site trackers. The browser keeps track of ad clicks that result in a conversion and sends this data back to the publisher's website (and potentially also to the advertiser's website as well) – with a random delay between 24 and 48 hours to prevent tracking based on observing conversion times, and without any user identifiers.<sup>364</sup>
362. This proposal supports 'click-through attribution', which gives credit for a conversion to an ad if the user has previously clicked on the ad within some time period before the conversion event. Safari does not appear to have any proposals to support 'view-through attribution', which gives credit to an ad if the person had been exposed to the ad before the conversion event.

#### *Chromium Privacy Sandbox – Click Through Conversion Measurement Event-Level API and Aggregated Reporting API*

363. In its Chromium 'Privacy Sandbox', Google has also proposed a new on-device technology for anonymous attribution, called Click Through Conversion Measurement Event-Level API.<sup>365</sup> Under this proposal, advertisers would be able to attach a set of metadata to their ads, which would be stored by the user's browser when the ad is clicked (including an impression ID, intended conversion destination, expiry dates). Similar to the WebKit approach, once

---

<sup>360</sup> See the [Mozilla blog announcing 'Today's Firefox Blocks Third-Party Tracking Cookies and Cryptomining by Default'](#).

<sup>361</sup> Indeed, Safari goes beyond making this a 'default' option as there is only one setting, which is to 'prevent cross-site tracking'.

<sup>362</sup> See WebKit blog, 'Privacy Preserving Ad Click Attribution For the Web', available [here](#). 'Online ads and the measurement of their effectiveness do not require Site A, where you clicked an ad, to learn that you purchased something on Site B. The only data needed for measurement is that *someone* who clicked an ad on Site A made a purchase on Site B.'

<sup>363</sup> Six bits can effectively encode 64 (2<sup>6</sup>) distinct values, and 12 bits means 4,096 (2<sup>12</sup>) distinct values.

<sup>364</sup> WebKit's draft specification for its Private Click Measurement proposal is available [here](#).

<sup>365</sup> See the [Click Through Conversion Measurement Event-Level API Explainer](#).



the user visits the intended destination page and converts, the browser records the conversion event and, some time later, sends a report to the publisher and advertiser (potentially via the a common adtech intermediary, such as an ad network) that a conversion occurred which can be attributed to a click on an impression, without the inclusion of any information by the user.

364. An important difference between Apple's proposal and Google's proposal is the size of the identifiers that can be used by advertisers to identify and disambiguate their ad impressions. WebKit suggests a very small 6-bit campaign ID, which effectively allows 64 distinct values to be stored. Google's proposal allows impression IDs up to 64 bits (ie  $2^{64}$ , or over 18 quintillion, distinct values), enough to uniquely identify every click. To compensate, Google's proposal involves sharply limiting the conversion data to just 3 bits (or 8 distinct values). Nevertheless, compared with Apple's proposal, Google's proposal allows for browsers to transfer significantly more information, and we have heard concerns that this potentially allows for more fine-grained mapping between impressions and conversions and a more significant risk of tracking of individual users.<sup>366</sup>
365. As with the WebKit proposal, Google's proposal is only well-developed for click-through attribution. Google is also working on extensions to this proposal which could support privacy-preserving view-through attribution and multi-touch attribution by integrating it with an aggregation service.<sup>367</sup>
366. Google also has a complementary proposal, called Aggregated Reporting API, which allows information from multiple websites to be collapsed into a single, privacy preserving report,<sup>368</sup> which could enable reporting of total ad views and reach for campaigns. The key idea is to make use of a write-only per-origin data store that reports data only if it reaches a sufficient aggregation threshold across many users.

## ***Practicability of privacy-enhancing technologies and proposals***

### *Technical challenges*

367. Privacy-enhancing technologies are the focus of a significant and ever-increasing body of academic literature. Progress in this area can open up the possibility of performing an increasing variety of common tasks (such as training a machine learning model) in ways that do not require direct,

---

<sup>366</sup> See EFF, '[Don't Play in Google's Privacy Sandbox](#)', section on conversion measurement.

<sup>367</sup> See [Conversion Measurement with Aggregation Explainer](#) and the [Multi-Browser Aggregation Service Explainer](#).

<sup>368</sup> See [Aggregate Reporting API proposal](#).

centralised access to data.<sup>369</sup> As an example, the area of Federated Learning has received increasing attention by researchers and practitioners alike.

368. Rapid future advances in these technologies might have the potential to preserve the efficiency advantages of the current digital advertising ecosystem, while tackling pervasive privacy issues. Since the Interim Report, we have seen more development of various proposals to apply these techniques to web browsing and digital advertising. It is encouraging that Google has recently started inviting adtech providers to test Privacy Sandbox proposals.<sup>370</sup>
369. Nevertheless, it is by no means certain that these proposals will succeed. There is a possibility that Google would reverse its decision to deprecate third-party cookies in Chrome if the Privacy Sandbox proposals were deemed insufficient or infeasible. Therefore, despite the promising progress in relevant technologies and privacy-enhancing proposals, the technological solutions that underpin many of these approaches are still in the development stage, and there remains uncertainty about their readiness. This point was made also by multiple stakeholders in response to our Interim Report, such as DMG Media, Verizon, Facebook, and the Developers Alliance.

#### *Commercial viability and regulatory support for adoption*

370. There is currently an active debate within the web standards community about restricting cross-site tracking. There appears to be significant momentum behind these efforts, following unilateral decisions by Apple and Mozilla to take stronger steps to implement browser changes to protect their users' privacy, and Google's announcement setting a two-year time frame to end support for third-party cookies. There appears to be at least a realistic prospect that the web standards community will achieve enough coordination around new privacy-enhancing standards.
371. In particular, the role of market-leading browsers in imposing standards, privacy-friendly defaults, and implementing privacy-enhancing technologies has been significant in encouraging websites to comply with standards. However, browsers still operate within limits and must balance their market position against those of important websites and web services. If significant websites and services do not support privacy-enhancing standards or technologies of certain browsers, and this results in a compromised

---

<sup>369</sup> See the previously cited Royal Society (2019) report for additional methods and applications of privacy-enhancing technologies.

<sup>370</sup> Digiday, '[Google is auditioning candidates to succeed the third-party cookie](#)', 13 May 2020.

experience to users, users may choose to stop using the web service or they may choose to switch browsers.

372. With respect to Privacy Sandbox proposals, Google plans to add each proposed new technology to Chromium once the development process is complete. Chromium is an open source project that provides the code used by many browsers including Microsoft Edge. Developers of these browsers will be able to implement the new technologies once available in Chromium, but there is no requirement that they do so; some may make modifications, and others may not use the updated code at all. Adoption depends on the web standards community. Similarly, with respect to ending support for third-party cookies, browser developers have the option to implement the new limitations in their own Chromium-based browsers but may decide to take a different approach.
373. There remains an important role for regulation and enforcement to support the efforts discussed in this section, and to guard against any tendency for certain stakeholders to drag out or unduly delay the standardisation process.<sup>371</sup> In any event, these technologies may be technically complex and costly to implement, and potentially require highly specialised talent to develop and maintain. This may restrict the number of entities that can effectively implement these solutions, and also raises the need for appropriate enforcement of data protection legislation in order to create the correct incentives to do so.
374. This view was supported by multiple stakeholders in response to our Interim Report. Various parties have suggested that a regulatory framework will be necessary to achieve adoption of these remedies at scale (CDEI, Oracle). Many parties deem necessary some standard-setting and regulation around design and implementation of PETs, if adoption is to be encouraged (DMG Media). The importance of ancillary measures, like operational / functional separation and a ban on cross-website tracking (as suggested in the Interim Report), is highlighted by Brave, the CLF, and the Horizon Research Institute.
375. Browser-based implementations of privacy-enhancing technologies could, in principle, cover both desktop and mobile devices with minor modifications. However, a large share of web traffic and advertising on mobile moves through apps rather than browsers directly. This might add a further obstacle to widespread adoption of privacy-enhancing technologies. It is notable that whilst Apple has taken a strong stance on the use of third-party cookies within the Safari browser, relatively little attention has been paid to the equivalent

---

<sup>371</sup> The failed experience of the [Do Not Track](#) initiative is instructive in this regard.

role of MAIDs within the app ecosystem. This will be an important area of future work, including for the recommended DMU.

376. As discussed in Chapter 10, we recognise the importance of ensuring a coordinated and coherent approach with other relevant regulators (such as the ICO), to prevent as far as possible duplication and inconsistencies between privacy, competition and other digital regulation. This point was made by many parties in response to our Interim Report (IAB, Facebook, Google, Advertising Association).

### ***Risks of privacy-enhancing technologies and potential for creating new concerns***

#### *Effect on users, publishers, and advertisers*

377. Most of the proposals available so far reduce the amount of user data that is exchanged. While this might alleviate privacy concerns, it might also have efficiency costs – a trade-off that is to some extent ineliminable. Coarser user data available to publishers and advertisers might make targeting and attribution efforts less precise.
378. Firstly, users might end up being exposed to somewhat less relevant ads. As far as users value ads that correspond to their interest, a less precise targeting would decrease welfare from their point of view.
379. There is a risk that, by reducing advertisers' targeting capabilities, publishers might incur significant revenue losses, jeopardising ad-supported models. As discussed in Appendix F, estimates of the value of behavioural targeted advertising for publishers vary widely, but can be quite substantial.<sup>372,373</sup> In a system where adoption of PETs is widespread, some of market participants' ability to conduct behavioural targeting would be retained, potentially

---

<sup>372</sup> See for example Johnson et al. (2017), [Consumer Privacy Choice in Online Advertising: Who Optes Out and at What Cost to Industry?](#), Simon Business School Working Paper No. FR 17-19; Marotta et al. (2019), [Online Tracking and Publishers' Revenues: An Empirical Analysis](#), Working paper.

<sup>373</sup> As discussed in Appendix F, current estimates of the value of behaviourally targeted advertising using advertiser bids and prices implicitly compare it to the value of advertising inventory with no associated cookie information or user profile, where both kinds of advertising (behavioural and contextual targeting) are available. However, in a counterfactual where behavioural advertising was prohibited, it is likely that some advertiser spending on behavioural targeted ads would divert to contextual advertising, rather than simply exiting the market. Therefore, publishers' revenues would not decline by the full value of the difference in the estimates of the value of behaviourally targeted advertising relative to non-behavioural targeted advertising. In addition, although advertisers may be willing to pay higher prices for inventory with richer user data that allows better behavioural targeting, fewer advertisers may compete for or be interested in reaching increasingly narrower consumer segments. The reduction in competition from fewer advertisers interested in each consumer segment might be sufficient to actually result in a net reduction in price of inventory, leading to a reduction in publishers' revenues. This point is made by Leven and Milgrom (2010). Levin, Jonathan and Paul Milgrom. (2010). [Online Advertising: Heterogeneity and Conflation in Market Design](#), American Economic Review: Papers & Proceedings, 100 (2), 603-607.

mitigating the impact on publisher revenues relative to an outright prohibition of behaviourally targeted advertising.

380. In response to our Interim Report, many news publishers expressed concern that the coarser targeting that PETs would imply might make digital news provision unsustainable (News Media Association and its affiliates), and called for us to balance the privacy benefits of PETs with their potential impact on the value of ad inventory.
381. As with reductions in targeting capability, current privacy-enhancing approaches will tend to impact the granularity and frequency of attribution data. This might reduce the efficiency of advertising campaigns, particularly those relying on real-time streams of ad click and conversion data.

#### *Effects on market power of existing platforms*

382. As discussed in the sections above on 'Relative value of and access to data' and 'Control over web standards and relative dependence on cookies', PETs potentially present significant implications for market power in digital advertising. Large, incumbent platforms have access to vast amounts of user data, obtained directly via their user-facing services. In a world where user data cannot be exchanged via cookies, and users cannot be openly tracked in their browsing activities, vertically integrated platforms with many users logged in to their services would still be able to deploy the granular data in their possession. This could potentially allow them to replicate many of the current targeting and attribution practices, while smaller non-integrated competitors would risk being foreclosed.
383. Similarly, large platforms have access to vast historical data on user behaviour and interactions with devices. Even if they were prevented from vertically sharing data from other services to their advertising arms, they would still have an advantage in the amount of data at their disposal for developing privacy-enhancing models. For example, in a world with FLoCs and no third-party cookies, platforms with leading user-facing services will still continue to receive more opportunities to observe user behaviour and which flock they belong to, and therefore more opportunities to make associations and draw inferences about flocks' aggregate interests and conversion behaviours, just as they currently have more opportunities to track individual users. Google and Facebook would therefore be better positioned to understand how to use flocks to target advertising. It is not clear whether market participants and adtech intermediaries that currently rely on the insights from third-party data about users, obtained from data brokers and

DMP marketplaces, will be able to replace this insight with information about flocks.<sup>374</sup>

384. As previously mentioned, successful application of PETs requires a shift towards on-device computation. Effectiveness and user experience are likely to be enhanced when these software technologies are seamlessly integrated with device hardware, especially in mobile.<sup>375</sup> Integration between software and hardware is managed by operating systems. Vertical platforms own most operating systems, especially in the mobile arena.
385. Multiple stakeholders expressed concern that client-side PETs, implemented on-device, would provide significant advantages to device and browser makers, who might have incentives to leverage their new gatekeeper role. This concern was echoed in different forms by the IAB, Prof. Geradin and Katsifis, and affiliate marketing operator AWIN, among others. For example, a solution like TURTLEDOVE would give Chrome a key role in advertising auctions, at least in the case of retargeting and possibly in all intermediated display advertising. Chrome would be responsible for selecting the intermediaries to which bid requests are sent and for executing the auction, functions that are currently undertaken by the publisher ad server. In the absence of common standards on how auctions are run, an intermediary, like Google, that also operates one of the most widely used browsers would have the ability (and possibly also the incentive) to favour its own intermediation services.<sup>376</sup> (Other examples of these kinds of conflicts of interest and leveraging market power concerns are discussed in Appendix M.)
386. In a situation where PETs were mandated as a standard, this might create an incentive for large platforms to provide privileged access to a device's computational resources to their own privacy-enhancing technology option, thereby creating barriers to new innovative entrants.
387. Another potential source of advantage for large platforms stems from the technical complexity of privacy-enhancing approaches. The development of such solutions is likely to require highly skilled computer science and engineering talent, with compensation levels that are almost exclusive to large tech firms.
388. As a general concern, several stakeholders raised the point that the strong market position of Chromium browsers may give rise to incentives for Google

---

<sup>374</sup> In this scenario, it is plausible that data brokers and DMPs would transition to buying and selling information about flocks, rather than for individual users or devices.

<sup>375</sup> For example, advanced federated learning application for mobile vision are only offered by Google on their own Pixel line of mobile devices – see [AI Google website](#).

<sup>376</sup> A key feature of Criteo's competing SPARROW proposal (discussed above) is that these key functions would be performed by an independent Gatekeeper, rather than by browsers that may be owned by adtech providers like Google.

(which dominates Chromium development) to build features that preference Google services. This concern was raised chiefly by 51Degrees, but the potential for such 'backdoors' was echoed by DMG Media and Prof Geradin and Katsifis.

389. These stakeholders also raised the concern that, despite an appearance of engagement with the web standards community, it's not clear that Google has a strong incentive to listen closely to other members of the community (adtech providers, advertisers, publishers), and might end up pushing through its favoured measures regardless. In this regard, 51Degrees also highlighted Google's extensive leadership and funding role in standard-setting internet governance bodies such as the W3C.

### **Further work with the ICO**

390. This appendix has identified various potential consumer protection and data protection concerns associated with user tracking. Some of these may warrant further work in collaboration with the ICO. In particular, we highlight the following issues:
- (a) mobile advertising IDs (MAIDs) are identifiers on mobile that are stronger than browser cookies but have received less attention;
  - (b) the permissions models on iOS and Android may facilitate inter-app data sharing beyond a user's awareness;
  - (c) the pre-installed apps ecosystem on the open source Android, means OEMs may install software with privileged access to consumer data without providing the user with any option to opt out, and scrutiny is impaired by certificate provenance issues;
  - (d) third party libraries (on both web and mobile) that websites or apps include offer extra functionality but also are usually not known to the user, and they may collect user data via the app/site;
  - (e) there is a case for monitoring and studying the consumer, competition and data protection impacts of browsers deprecating third-party cookies and alternative approaches to targeting and attribution;
  - (f) assessments should also be made of the case for supporting the development of privacy-enhancing technologies (PETS); and
  - (g) engagement with internet governance forums may be useful, including browsers and standards bodies such as the IETF, W3C and WHATWG where normative technical standards are developed internet-wide.



391. We will draw on this list of issues in taking forward the joint work with the ICO set out in Chapter 10.