Legislative Council Staff
Nonpartisan Services for Colorado's Legislature

Memorandum

March 2026

TO: Interested Persons

FROM: Dhivahari Vivek, Science and Technology Policy Program Fellow

SUBJECT: Online Data Collection and Tracking Methods by Third Parties

This memorandum provides an overview of known online data collection and tracking methods often used by third parties, including data brokers. While the ecosystem of online data collection and tracking is vast, this memorandum primarily focuses on user side collection and tracking methods and briefly describes ongoing research regarding non-user-side third-party data tracking and brokerage.

Re-Identifying "Anonymized" Data

Personal data varies in its identifiability. For example, data can directly identify an individual (e.g., name, phone number, or SSN). In addition, indirect identifiers, such as gender, date of birth, or zip code, can, in combination, be unambiguously linked to an individual. One study was able to uniquely identify at least 63 percent of the U.S. population by the combination of indirect identifiers (e.g., zip code) available from census data (Lubarsky 2017). Below this level of identifiability rests data that can be ambiguously connected to multiple people, like physical measurements or restaurant preferences.

Much of the data collected by the methods discussed in this memo may be considered "anonymized" data—data scrubbed of certain types of personally identifiable information (PII) such as a name or Social Security number (SSN). However, the proliferation of both publicly available information online and data collected by the online tracking methods discussed below, combined with increasingly powerful computer hardware, make it possible to re-identify "anonymized" data. Studies involving location cell phone data have demonstrated that as few as four pieces of spatiotemporal data points (e.g., precise location coordinates and a timestamp) could uniquely identify as much as 95 percent of the population.

The most powerful tool for re-identifying scrubbed "anonymized" data is by combing multiple datasets that contain the same individual(s) in each dataset. Processes like "pseudonymization," in which direct identifiers are replaced with unique random identifiers and are frequently employed by data brokering companies, can be reversed to re-identify individuals in a number of ways (Lubarsky 2017). The datasets built from the data collection and tracking tools described in this memo are frequently processed and combined to re-identify and establish comprehensive user profiles of individuals.

Connecting to the World Wide Web

When a user visits a website, the web server hosting the site requires basic information including the visiting user's Internet Protocol (IP) address to establish a connection between the web server and the user's computer and to load the web page. Other required functional elements of the website connection process collect data like the user's browser type and version, operating system (OS), and language preferences. This type of data collection and web tracking is typical and generally required for a website to execute basic functions, such as adapting content for mobile versus desktop viewing or checking browser compatibility.

Online persistent tracking is done by linking specific devices or users to a **unique identifier**. There are several types of unique identifiers:

- **Hardware device identifiers**. For example, Media Access Control (MAC) addresses are a sequence of twelve hexadecimal digits (0-9 and A-F) that identifies any device connected to a network. This identifier is linked to the physical device hardware, assigned by device manufacturers, and unique to each device. Mobile devices can also have hardware identifiers, like the International Mobile Equipment Identity (IMEI).
- **Web-based identifiers**. IP addresses are virtual numerical labels assigned by a router to each device connected to a computer network that use the IP for communication. Whereas MAC addresses identify a device within a network, IP addresses are used to tell devices and routers where to send data. IP addresses serve two main functions: identifying a specific device on a network, and locating the device's position within a network to enable communication with other devices on that network and across networks. IP addresses can change for a number of reasons, including when a user's device connects to a new network (e.g., the IP address of a device connected to a home network versus a coffee shop will be different), upon restart of a router, or to improve privacy (e.g., by using a VPN). Additionally, routers assign private IP addresses for device communication within a local network (such as a home network) and public IP addresses when a device connects to the internet.

- **Mobile device identifiers**. Mobile devices can also be linked to mobile advertising IDs (MAIDs). MAIDs are unique identifiers assigned by the device's OS (e.g. iOS, Android) to mobile devices to help advertisers track and personalize ads. MAIDs, which comprise a long string of numbers and letters, enable targeted ads to be sent to a specific device based on data collected about the user, including app usage and behavior and ad interactions to allow for ad retargeting. Apple's iPhone OS (iOS) calls this the "Identifier for Advertisers" (IDFA), and on devices running iOS and iPadOS 14.5 or later, Apple requires developers to explicitly request users to authorize tracking. Google's mobile OS (Android) calls this the "Advertising ID;" in late 2021, Google Play services started removing the advertising ID when a user deletes their advertising ID in Android settings.
- **Third party identifiers**. Third parties may also assign their own identifiers to track user behavior, within and across sites, using a number of tools discussed below.

Data Collection and Tracking Tools

There are a number of tools and methods used to retrieve user data, including many used by third parties. Below is an overview of key user-side data collection tools commonly used by website and app operators.

Website Cookies

Used by website operators/publishers and third parties, cookies are small data files that are often a fundamental element of website applications. When a user visits a website, cookies are created and stored locally within a browser's folders on the user's hard drive. Cookies can store unique identifiers that can track users across multiple visits to the same site, or across visits to different sites. Cookies can serve the following functions:

- session cookies perform essential site functions;
- analytics/performance cookies track a user's site behavior to evaluate and improve site performance; and
- advertising cookies track a user's online behavior to enable targeted advertising.

Temporary cookies, like single-session cookies, only last for the duration of a single visit to a website, while persistent cookies can remain on a user's hard drive long after an initial website visit and continuously record user website behavior, until they are manually deleted by the user or expire (Cahn et al. 2016).

First-party cookies

First-party cookies are created by the website visited by the user and often include "essential" cookies. They enable same-site tracking, and are commonly used for basic site functionality and collecting analytics about user engagement that can be helpful to online businesses. First-party cookies help enable features like remembering the contents of a user's shopping cart across multiple visits to the same e-commerce website, browsing history, site preferences, or saving a user's preferred language setting. Websites may place cookies on the user's web browser so that their device is recognized if the user visits the same website in the future.

Third-party cookies

Also known as tracking cookies or cross-site tracking cookies, these are created by a third-party entity other than the operator or publisher of the website visited by the user. These cookies enable advertisers and websites to track user behavior and build user profiles to serve personalized ads. Third-party cookies can be sent to a user if that user clicks on a website link that redirects to a third-party site, or if a webpage embeds components from other sites (like third-party ads or social media buttons such as "Like" and "Share").

Third-party cookies are routinely deployed by data brokers, online advertisers, third-party advertising networks (businesses that connect advertisers with websites or apps that want to show ads), and tracking applications (e.g. Google Analytics) (Cahn et al. 2016). If a user visits multiple websites that uses the same third-party service, that third party can identify that user's tracking cookie with its unique identifier across those sites, recognize that the same browser or device is visiting a new site, and allow that third-party service to link these various sites to a particular user, creating a profile of that user's browsing habits (Nikiforakis et al. 2013). In situations where a website serves ads from different ad networks, multiple third parties can link their own cookie identifiers to recognize the same user – a process known as cookie syncing, which a majority of third-party cookies utilize.

Device and Browser Fingerprinting

Fingerprinting algorithms combine data collected to create a "fingerprint," including data about the device or browser properties (browser settings and version), information about the system's video/audio processing capabilities, and computer display size and resolution. Device identification can be performed through web-based fingerprinting, where a website identifies a particular browser (and users by extension) by collecting and combining distinguishing features of the browser and underlying operating system. In fingerprinting, a piece of code on a website typically collects device and browser data. While fingerprinting is often used for constructive means such as combating fraud, fingerprinting can also be used to track users between sites. Unlike cookies, a device or browser's "fingerprint" generally lasts as long as a user has the same hardware and software. Existing research implies that the majority of users' browsers are uniquely identifiable (Cho et al. 2022).

Tracking Pixels

Tracking pixels are small pieces of code used for online advertising, and are embedded in websites, ad banners, and emails to collect and send data to a third party. This code creates an image the size of a single pixel (not visible to users) that is triggered to send data to the pixel provider's servers when a user visits a web page (Cho et al. 2022). Tracking pixels today have also evolved from pixel-sized images to include a broad range of HTML and JavaScript code embedded in websites and email – making the term "pixel" sometimes misleading.

In addition to device and browser information, data collected by tracking pixels also includes user online behavior, such as how a user interacts with a web page, including specific items a user has purchased, or information that users have typed within a form while on the site. One commonly used pixel is the Facebook/Meta Pixel (offered by Meta), and is used by more than 30 percent of popular websites with at least 2.2 million Facebook pixels installed on websites by June 2018.

Popular tracking pixels can also rely on other tracking tools, mainly cookies. For example:

- For websites that embed the **Meta Pixel**, the pixel uses Facebook cookies to match a website's visitors to their respective Facebook user accounts and record the user's actions in their Facebook Ads Manager. In cases where cookies are not enough to match a user's browsing behavior to a Facebook or Instagram account, Meta also allows the website to send user information through other means, such as from information submitted in a website form by the user.
- When using the **TikTok Pixel**, third-party cookies are on by default, and enhances the pixel's ad optimization abilities, like measuring conversions (e.g., tracking website visits from ad views on the TikTok app).

Application Programming Interfaces (APIs)

APIs act as intermediaries that enable digital platforms and services to interact with one another and share data, much like how two people who want to send letters to each other would need an intermediary (such as a postal service). In the same way that a postal service has predefined rules for someone who might need to send or receive a letter (e.g., addressing format), APIs create a set of allowed interactions, rules for achieving those interactions, and a service address that can receive and handle data requests. Websites and apps can follow those rules to share information with each other using the API.

Both third-party developers and advertisers use APIs to share and retrieve data from online platforms, and not all APIs directly share insightful user data with developers (Russell et al 2019). Certain types of APIs can facilitate user data sharing between developers and large technology platforms. Most large tech companies offer APIs that allow independent websites or mobile apps to integrate their services – one type of API known as **feature APIs**. For example, Meta offers the Facebook Login API that independent developers can incorporate to enable users to log into their website or app using their Facebook account, rather than requiring users to create a separate account. These types of APIs can provide convenience for both the user (who does not have to create and remember login credentials for a new online account) and for the developer (who does not have to write original code to ask for and store user login credentials). However, the independent website or app will gain access to a user's Facebook account, including a person's real name and profile picture if that user's profile is public.[1]

Many websites use APIs from large tech companies, which may collect and track users across different websites that use their APIs. While the extent of detail learned about the user may vary, at a minimum, companies that offer certain types of APIs are able to identify users. This generally occurs in situations where, for example, a user interacts with a Google, Apple, or Facebook feature API while being logged into a Google, Apple, or Meta account on the same device or web browser. Other tracking tools like cookies can augment this tracking process, enabling large tech companies to consistently identify users on any website that uses their feature API. This tracking can be done even if the user does not actually use the feature API. If, for example, a user decides instead to create a new username and password to access a website that offers Facebook Login – and is logged into a Facebook account on that web browser – the existence of the code on that website that enables the Facebook Login API could share with Meta what users have visited that website (Russell et al. 2019).

Other types of APIs that may connect app developers to user data include:

[1] Meta for Developers, Facebook Login Overview. Companies like Meta, Amazon, and Google now require apps to go through a review process if developers request user permission to access more granular data. For example, Meta cites selecting unneeded permissions as a common reason for rejection during app review.

- **Analytics APIs**, which allow third-party services to gain information about the visitors to their website.
- **Unofficial APIs**, such as when a developer examines and mimics the way a legitimate mobile app internally communicates with the accompanying legitimate app servers. Users may encounter these APIs if they use a third-party application to sign into a legitimate service. For example, if a user connects their Venmo account to a budgeting app, that budgeting app may be able to access (and feasibly store and process) that user's Venmo account details and transaction history.

APIs can also play a critical role in defining advertiser-consumer relationships through **advertising APIs.** Programmatic advertising, which involves a real-time bidding process and is discussed further below, makes use of advertising APIs.

Software Development Kits (SDKs)

SDKs are a set of tools provided generally by the manufacturer of a hardware platform, OS, or programming language, intended to help software developers create applications for that specific hardware platform, OS, or programming language. SDKs are intended to aid developers when building apps for specific platforms and integrating their apps with a vendor's service. Instead of writing code for critical app functions purely from scratch, developers can license an SDK's many helpful tools including APIs, reusable code to perform specific functions, visual editors to help with graphic design, testing and debugging tools, and documentation with clear instructions throughout app development. Google and Meta are two of the most popular mobile SDK providers (Cho et al. 2022). Some SDKs, like ones required to ensure a mobile app's compatibility with an OS, are not intended for mobile tracking. SDKs can make mobile app development faster and easier, and subsequently are commonly used.

Some SDKs, including third-party SDKs, specifically track and collect user data across mobile apps. App developers may use advertising or ad network SDKs to obtain revenue (such as Meta's advertising SDK). These types of SDKs enable apps to send targeted ads to users based on the user's interaction with the app, and often in conjunction with other data collected by the SDK supplier. Consumers are often unaware of the integration of third-party SDKs in mobile apps, and may unknowingly share many kinds of data with third-party SDKs when interacting with a mobile app. Even app developers themselves may be unaware of the extent of data collection by third-party SDKs (Cho et al. 2022).

Emerging Research on Third-Party Tracking

Much about how third parties collect consumer data is still unknown and requires further research. This includes methods of tracking that are not user-side methods. Ongoing or under-explored areas of research regarding online consumer data collection and tracking include the following.

- **Server-side tracking (SST).** Rather than tracking performed on the client side (for example, on the end user's device) using the many methods described above, a web browser may instead forward data to an intermediary server. For example, in client-side tracking featuring cookies (as described above), a website visited by the user may include content from third party sites looking to track users – and the browser makes requests to these third-party sites to retrieve this content, and, as part of this process, stores its third-party tracking cookies. However, in SST, user data is instead sent to a private server controlled by the original visiting website (called a first-party subdomain), which then forwards user data to third-party sites and advertising vendors. The data collection that happens through SST is made to look like the user's browser is interacting with the original visited website, masking the browser's tracking by third-party sites (Fouad et al. 2024). Mechanisms like CNAME cloaking – where a third-party website's interactions can appear to a user's browser as if it came from the intended visiting website – can also be leveraged to pass third-party content (like cookies) as first-party.
- **Tracking by mobile cell carrier networks.** Cell carriers can track device location by devices simply connecting to cell towers. Each time a device a.) interacts with a cell network (making a phone call, sending a text message, or using cellular data), b.) transfers connection between cell towers, or c.) goes unobserved for a set period, a data point is created, which can include data like a unique identifier (such as a subscriber ID and hardware identifiers), a timestamp, the device's precise coordinates, and the nationality of the subscriber's home network (Doorley et al. 2020). Mobile carriers may combine location information with collected third-party data for advertising purposes. Verizon collects cell tower location information among other types of user information, and also collects user information from third-party data sources. Verizon may use this collected user information to more accurately predict and deliver personalized content, including ads for Verizon products or third-party products and services sold by Verizon, that may be placed based on geographic information.[2] Further research is needed regarding the possible role of mobile cell carriers in

[2] As of March 3, 2026, information regarding Verizon's data collection practices can be found by accessing Verizon's Privacy Policy, under "What information does Verizon collect?". This section discloses that

the data brokerage ecosystem. As an example, one study estimating the impact of the COVID-19 pandemic on population changes in Andorra leveraged mobile carrier data to measure user mobility within the region (Berke et al. 2022). These methods may be repurposed to track users for advertising purposes. In the United States, federal law requires carriers to take reasonable measures to protect certain customer information, including location information.[3] In 2024, the FCC fined the nation's largest wireless carriers for illegally selling location data access to third parties without express customer consent and for not taking reasonable measures to protect that data against unauthorized disclosure.

- **Advertising exchanges and programmatic advertising.** Recent reporting and cases pursued by the Federal Trade Commission (FTC) have focused on how data brokers can obtain user data through advertising exchanges. These are digital marketplaces where publishers (e.g., website and app owners) auction ad space on their platforms to advertisers in a process known as real-time bidding (RTB). When a user visits a website or opens an app partnered with an advertising exchange, their data is shared with the ad exchange (e.g., IP addresses, MAIDs, device hardware information, precise GPS coordinates of the device if the user enabled location sharing on their device, browsing behavior, and timestamp of data collection). This data is sent by publishers as a "bid request." This request is then broadcast to potential advertisers bidding on the spot, bid responses are evaluated under prevailing auction rules, and a winner is selected. The entire process takes place within milliseconds. All bidders, even losing bidders, get access to the broadcasted user data during the RTB process. The granular level of detail collected enables advertisers to serve ads directly to consumers based on their demographic, psychographic, or behavioral attributes. Big platforms like Meta and Google usually operate their own RTB marketplaces, while relatively smaller companies rely on third-party exchanges (Russell et al., 2019).

Verizon obtains information from outside sources, including third-party data providers, and that Verizon uses cell tower location information under "Location of your wireless devices". Information pertaining to how this information can be used is found in the section titled "How does Verizon use information?", under "Better predict and deliver personalized content and offers that may interest you".

[3] 47 U.S.C, § 222