

Mr. Simeon Thornton
Project Director, Mobile Ecosystems
Competition and Markets Authority
The Cabot, 25 Cabot Square
London, E14 4QZ
United Kingdom

Oracle Response to Mobile Ecosystems Market Study

Computing is now mobile. What was once an industry and a marketplace confined to bulky, and sometimes hulking desktops, has been transformed by mobility. The ability to access and transmit data from anywhere has revolutionized work, shopping, and communications. Mobile computing made businesses more efficient and has enabled entirely new conveniences in the consumer economy that are now commonplace. Highlighted by a year of lockdowns and quarantines, our mobile devices and the networks we depend on kept families connected, businesses operating, and ensured movement of essential services and materiel.

While the mobile ecosystem has become essential for every business and every consumer in our online world, there exists a serious competition problem at its heart. The CMA's "Mobile Ecosystems Market Study" is an important undertaking, and one that Oracle welcomes. As the CMA correctly notes, mobile devices are dominated by one of two operating systems, both of which are important products of their two parent companies. As the CMA notes, Apple's iOS represents about 52% of the UK's market in mobile phone handsets with Alphabet's Android taking the other 48%. Yet it should be noted, as the CMA begins its study, that the mobile ecosystem is broader than smartphones and encompasses millions of internet-connected devices from smart speakers, to home appliances, to increasingly, automobiles. Taken as a whole, Android, because of its prominence among these devices, is the world's largest operating system.

There is also a fundamental difference at the center of the operating system duopoly. One system, iOS, is proprietary and works as Apple's user experience portal for its customers. Its role is to enable Apple's products to work in the expected manner across devices. Apple's profits come primarily from the sale of hardware and subscriptions. For Apple, the role of iOS is to build brand loyalty through an integrated product line. For Google, Android serves a much different function.

Android is a mechanism for data collection. It is licensed by Google to original equipment manufacturers (OEMs) and is integral to Google's core business model of selling advertising. As we summarize below, and have explained in previous submissions to the CMA, the Android operating system provides Google with a constant stream of data which, using a series of unique device identifiers, correlates the data with a particular consumer, device, or account. This correlation allows Google to create "super-profiles" of individual consumers from the streams of location, app usage, internet browsing, purchasing, and activity data. Most consumers only have a scant understanding of this collection. Opt-outs are by turns cumbersome, incomplete, or illusory. An Android user concerned about the amount of data

collected from the device has little control over the totality of data collected from them, especially if they expect to use their device as anything more than a cellular telephone.

Oracle has researched Android's data collection for close to six years, and in previous submissions to the CMA has provided detailed explanations on how Google uses Android's privileged position as the device operating system to siphon a huge variety and volume of consumer data to Google. The richness of these data sets is the catalyst for Google's dominance in advertising. Not only does Google have more data than anyone else about consumers, it also has that information in real-time – in particular the location and current activity of every Android phone on the globe. This creates a virtuous cycle for Google – more data brings more advertisers, which in turn attracts a galaxy of developers and publishers who provide Google with even more data, which Google then uses to hone their targeting.

No single company can hope to replicate this dominance, because in order to compete a new entrant would need all of the tools at Google's disposal – a search engine, a browser, an app store, and an operating system. These structural advantages are buttressed by a host of contracts and agreements with OEMs and carriers that effectively limit competition and mandate data sharing.

In the next section, we will summarize how Google uses Android to create these super-profiles and how the notice and consent framework underpinning the data collection is fundamentally flawed. It is our belief that in order to create more competition and privacy protection in these markets, the CMA must understand how these problematic data collection policies are arrayed against the consumer, and how this imbalance is fundamental to Google's data collection.

Google's Data Collection and Consents

In previous submissions for the CMA, Oracle has provided documentation of how Google uses Android to collect intimate personal and “pattern of life” data from users through an intertwined system of device functionalities and consents. The level of data collection is more invasive than Google's competitor in the market for mobile operating systems, Apple, and is certainly out of proportion with user expectations. Secondly, even for the small set of consumers who understand the level and granularity of data collection, opting out is difficult, presuming those mechanisms can even be found in the device's settings. In addition to our previous submissions to the CMA, we also recently [provided comments to the United States Federal Trade Commission](#) workshop on dark patterns, providing more detail about how Google manipulates users to maximize data collection from Android devices. We also provide an updated paper on how Google assembles Android data into “super-profiles”.

Google's properties collect user data across the internet, but none is more pervasive than the Android Operating System. In order to use a device running Android, a consumer must agree to Google's Terms of Service and Privacy Policy (Figure 1), which give Google permission to collect a vast amount of user data and can change at a moment's notice. Google makes it difficult, if not impossible, to opt-out or even limit data collection. Where Google does provide limited opt-out options, users are forced to sacrifice functionality to do so.

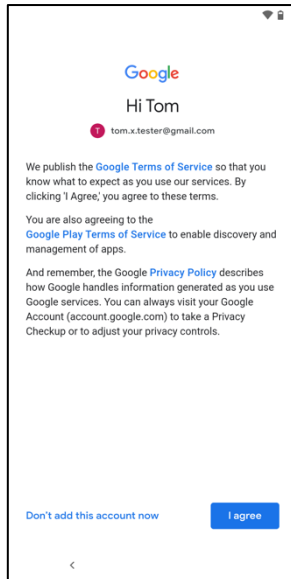


Figure 1: Google Terms of Service and Privacy Policy Agreement during Android Set-Up

When configuring an Android device, location is turned on by default as are a number of location related sub settings (Figure 2). “Wi-Fi Scanning” and “Bluetooth scanning,” are both turned on by default as is “Location Accuracy.” Together, Scanning and Location enable the collection of nearby Wi-Fi base stations and Bluetooth beacons, even when Wi-Fi or Bluetooth are disabled, and the device’s GPS coordinates.

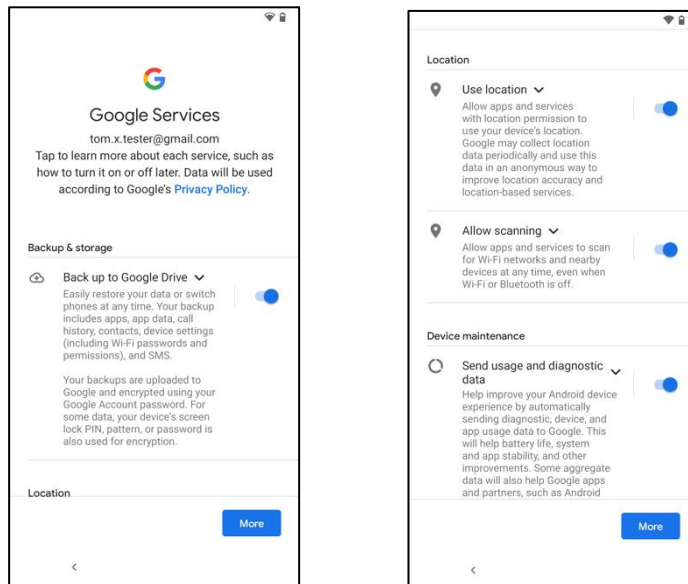


Figure 2: Default Android 11 Location Settings

Where Google does offer limited opt-out options for users, these choices require users to limit device functionality. To stop the collection of nearby Wi-Fi base stations and Bluetooth beacons, a user must both disable Scanning and turn off the Wi-Fi and Bluetooth settings. If done successfully, the user will not be able to use the Wi-Fi and Bluetooth functions of their device. Android’s newest location setting launched in August 2020 and is called “Earthquake Alerts¹.” This setting is turned on by default and allows Google to collect device sensor readings, such as barometer and accelerometer readings. Moreover, when the Earthquake Alerts setting is disabled, only user notification is turned off and Google will continue collecting the device’s sensor readings (figure 3). To disable Google’s collection of device sensor readings, the user must navigate to a completely different setting called “Location Accuracy” and turn it off (Figure 3). Turning off Location Accuracy limits functionality across the device. According to Google, “when you turn off Google Location Accuracy, your phone uses only GPS to find location. GPS can be slower and less accurate than other sources².” Location aware services such as Google Maps and third-party apps such as Uber do not work as well when Location Accuracy is turned off.

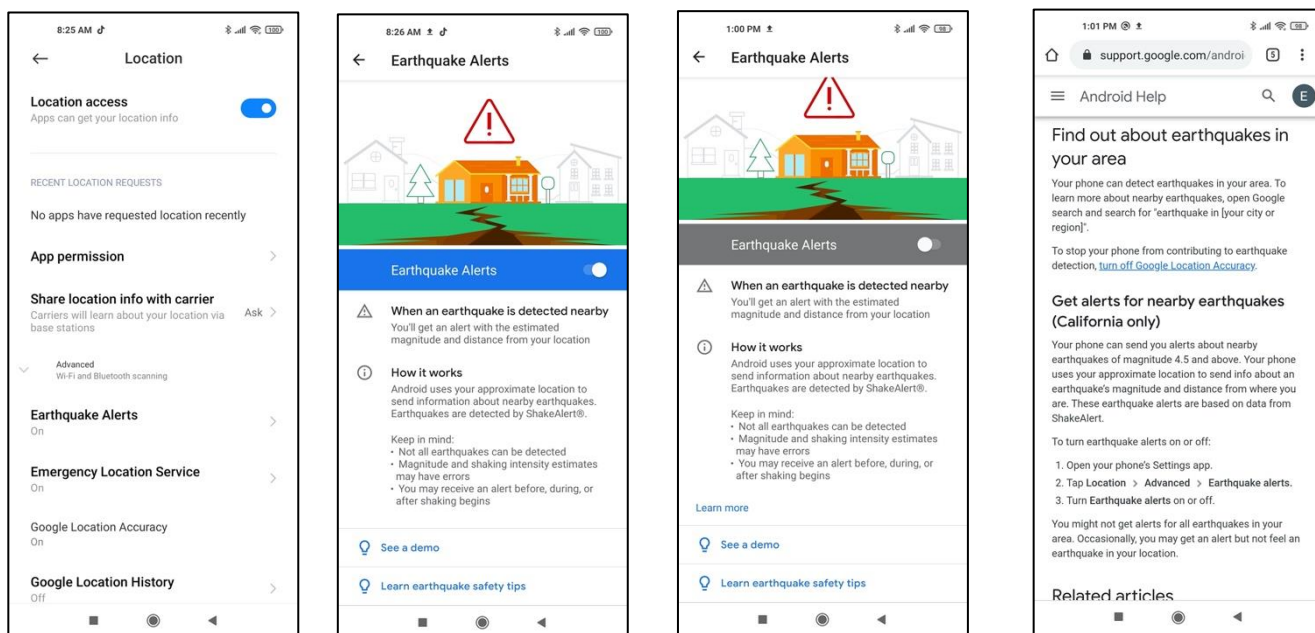


Figure 3: Android Earthquake Alerts Setting Screens

¹ <https://blog.google/products/android/introducing-android-earthquake-alerts-outside-us/>

² <https://support.google.com/nexus/answer/3467281?hl=en#zippy=%2Cwhen-google-location-accuracy-is-off>

Figure 4, below, provides a small sample of the cataloging of personal data that underpins the Android mobile ecosystem. The data covers almost the entirety of a user’s actions, locations, the wi-fi network the phone is near, search information, as well as a host of specifics about the device. As the device moves through the world, its contact with networks in proximity to it, provide a constantly updating stream of data about a consumer’s whereabouts and activities.

Data Screenshots

id info - id info page	20 unique found	+
ad info - ad info page	1 unique found	+
loc info - loc info page	106 locations referenced (9 unique)	+
place info - place info page	1 found	+
wifi info - wifi info page	12 unique macs in 113 scans at 105 locations	+
activ info - activ info page	106 loc/131 wifiscan/3 wifiConn/5 rate upload/9 activity	+
social info - social info page	1 found	+
intent info - intent info page	1 searches with 1 suggestions	+
phone info - phone info page	19 unique found	+

Figure 4: Sample Categories of Data Collected by Google from an Android Smartphone

loc info	
07-22 08:27:28 (EDT)	wifi 48.8819933 , 2.3630157
07-22 08:37:35 (EDT)	wifi 48.8819933 , 2.3630157
07-22 08:47:53 (EDT)	wifi 48.8819933 , 2.3630157
07-22 08:51:56 (EDT)	wifi 48.8819804 , 2.3630253
07-22 09:02:09 (EDT)	wifi 48.8819804 , 2.3630253
07-22 09:12:44 (EDT)	wifi 48.8819804 , 2.3630253
07-22 09:22:50 (EDT)	wifi 48.8819804 , 2.3630253

Figure 5: Sample Location Data Collected by Google from an Android Smartphone

sensor info - sensor info page	
08-07 16:18:09 (EDT)	loc 10 points
08-07 16:18:13 (EDT)	baro 218 readings
08-07 16:18:09 (EDT)	accel 749 readings
08-07 16:18:09 (EDT)	mag 2246 readings
08-07 16:18:09 (EDT)	gyro 2249 readings
08-07 16:19:47 (EDT)	baro 228 readings
08-07 16:19:42 (EDT)	accel 784 readings
08-07 16:19:42 (EDT)	gyro 2352 readings
08-07 16:19:42 (EDT)	mag 2352 readings
08-07 16:21:28 (EDT)	baro 250 readings
08-07 16:21:23 (EDT)	accel 856 readings
08-07 16:21:23 (EDT)	gyro 2570 readings
08-07 16:21:23 (EDT)	mag 2570 readings
08-07 16:22:23 (EDT)	baro 62 readings
08-07 16:22:22 (EDT)	accel 213 readings
08-07 16:22:22 (EDT)	mag 638 readings
08-07 16:22:22 (EDT)	gyro 639 readings

Figure 6: Sample Device Sensor Readings Collected by Google

Remedial Actions to Consider

The CMA will likely find, as was found in previous studies on digital advertising and online platforms, that the mobile ecosystem is likewise, not a well-functioning competitive market. The question then becomes, how to correct for the structural asymmetries and how to encourage competition. Oracle agrees with the conceptual framework categorizing corrective actions into four main themes outlined in the Statement of Scope.

The global mobile marketplace is dominated by two firms, with Android representing 73% of global operating system market share.³ Location independence is the central feature of mobile, being able to access information without regard to place defines the mobile experience. Without a common set of rules across countries, action in one area could lead to less regulation and more rent-seeking elsewhere.

Interventions Limiting Platform Power

Taken as a whole, each of these categories, seeks to limit platform power. For the purpose of this remedy, the CMA should examine the outsized importance of the operating system for the mobile device itself. An operating system is the connective tissue of the device, allowing various features of the phone to operate in concert with one another. When a consumer first chooses a phone, that choice is typically not driven by the features of the operating system, instead the choice is most likely driven by the features of the phone (for example, screen size, camera resolution, speed, battery life).

Much like an airplane or automobile, a smartphone is a product of the modern supply chain made up of chipset manufacturers, telecommunication network providers, device OEMs, app developers, and the operating system vendor. All of these functions must work together. Despite

³ <https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/>

this symbiotic environment, the lion's share of the wealth and market power created by the phone currently accrues to the provider of the operating system. This wealth transfer is a result of the data pouring off the handset and is guarded by restrictive contracts on the network providers and OEMs that restricts their ability to use the data to compete against the operating system. There is nothing in the engineering or design of mobile phones that predetermine which entity should derive the most benefit from their operation, the only factor directing this are Google's contractual restrictions that effectively limit competition.

Google is able to make these demands because consumers are largely locked into an operating system ecosystem from the moment they choose their first smartphone. The switching costs, in terms of the actual device and the accumulation of products, services, and consumer data on the device, quickly becomes too great. With a captive customer base created by customer lock-in, network effects, and exclusionary conducts, Google is able to extract concessions from developers, telecommunication carriers, and device manufacturers in order to provide the Android operating system. Compared to Apple iOS, the Android operating system has a robust data extraction function that delivers real time, highly precise location data as well as search and app usage data. This data is not shared with the network provider, the OEMs, or the app developers, all of whom are equally, if not more, important to the customer. Instead, the data is delivered to Google where it is collated and monetized for the provision of advertisements.

I. Interoperability and Common Standards

Oracle supports increased interoperability and portability to the mobile ecosystem. As we stated above, there is nothing inherent to an operating system that determines which entity can, or cannot, provide services to a consumer or receive data from a user. That reality is a creation of licensing agreements, mobile application development agreements, and a host of other contracts imposed by Google on handset suppliers and developers.

It is apparent at the outset that for a smartphone to function in the manner expected by a user it must interoperate. The problem, particularly with Android, is that Google has largely predetermined with whom, how, and who gets the most benefit from the interoperability. This puts Google in a curious position, offering some interoperability but clearly not enough to create additional competition in the marketplace.

Particularly in the context of mobile functionality, interoperability is linked to data portability – the ability for consumers to have other market participants access and provide services for their data. Developing a mobile ecosystem where interoperability and portability are central features, would spur competition and innovation. The current state of the market, where one company receives limitless data and where consumers have little say in the transaction, is unacceptable.

Developing common standards is a prerequisite for creating improvements in interoperability. Any choices provided to consumers must be seamless and any technical problem between service providers would inevitably result in the failure of a competing product. Standardization comes with a note of caution. Too often, dominant incumbents usurp discussion

of industry-wide standards. Efforts at creating standards for interoperability, including through mandatory disclosures of APIs, should guard against an incumbent simply recreating their current market advantages into the establishment of new standards.

II. Consumer Choice Remedies

Allowing consumers to determine how their own data is used, and by whom, might be the most important way to inject competition into this market. We believe this can be readily achieved. In fact, a useful analogue is already an expectation of consumers in the porting of phone numbers between network providers.

Mobile devices, particularly handsets, can transmit data to a variety of entities. Currently, most commercially valuable data is transmitted to the operating system provider. This data also often contains the most intimate and private information about the device user. Under current device and service agreements, the consumer has little recourse about who can see this data or how it is ultimately monetized. Using the unique identifiers on the device, Google is able to build a profile of each consumer, tying movements, searches, app usage, financial transaction, and pattern of life to the opaque string of numbers assigned to the device. Yet there is nothing inherently Google's about these identifiers.

Providing consumers with the choice about who, and for what purposes, those identifiers and the streams of data associated with them can be accessed is fundamentally important for a competitive mobile ecosystem. This one change can enable new entrants in all manner of services currently solely dominated by Google, or in tandem by the operating system duopoly. Consumer choice, particularly over who accesses the data stream, would mean a true "offer and acceptance" model and most likely one where new entrants would have to offer something of cognizable value to the consumer for their data. Currently the offer is a functioning phone in exchange for all of your data in perpetuity for the purposes of advertising, or any other future use – with the opportunity to revise the terms at any time for the express benefit of Google.

One effect of providing consumers with this enhanced choice would be inevitable competition for privacy and security. Data breaches could be met with loss of customers, where now consumers have little recourse to respond. Consumers concerned about the use of their personal data could decide for themselves to hold their data private, or entrust it with a market participant who has better privacy protection.

Providing consumers with these choices would also necessitate remaking the restrictive contracts Google has put into place on the thousands of entities that exist to give Android phones the rich environment to attract consumers. App developers, chip manufactures, OEMs, and network operators could all compete for consumer loyalty based on the quality of their offering, rather than have one dominant entity deny competition based on network effect and exclusionary conduct.

III. Structural Separation

Google's dominant position in the mobile ecosystem could be limited through simply separating it from Android. As Google itself repeatedly makes clear, Android is engineered to be an open source operating system. However, with the exception of the Chinese market, devices with open source Android are incredibly rare. That is because Google has used its market power to ensure that Android devices remain vassals to Google. Separation could at once create a more robust market, enhance competition, spur innovation, and provide consumers with more choice.

Android, if separated from Google, could be run as an independent entity allowing for non-discriminatory and open access for developers. We recommend a [recent paper by Yale University](#) on this subject.

Structural separation offers a series of advantages to policy makers and enforcers than the behavioral remedies, discussed above, do not. Simply put, a structural separation is easier to manage and more difficult for the target to manipulate. It would immediately create dividends in the marketplace and could allow for more competition than in an environment marked by active management from policy makers and enforcers.

Oracle appreciates the opportunity to offer these comments on this proceeding. We stand ready to answer any questions you may have.

Google's Shadow Profile: A Dossier of Consumers Online and Real World Life

June 2021

Executive Summary

A consumer sees an ad that is unnervingly, pointedly accurate. It seems to target information – so personal, so specific – that only this consumer would know the information. Maybe the ad targets a secret interest or hobby, a special place, or intimate lifestyle details. Is the microphone on? Is the camera activated? No –but they might as well be. In fact, Google is using massive amounts of consumer data, not all of which it discloses to consumers, to micro-targeting advertising. All without the consumers knowledge or consent.

Google's corporate mission is "to organize the world's information and make it universally accessible and useful." What it does not widely acknowledge is that this mission is as much about collecting data as it is about categorizing information. Google acknowledges certain data collection activities, and even purports to grant consumers control over what is collected and how it is used. However, the scope and extent of Google's data collection extends far beyond what is acknowledged or widely known, and its controls fail to address most of this data. As a result, consumers cannot fully understand – much less control – all of the data that Google holds on them.

While Google touts "improved" consumer control over the data it collects (as a result of the pressure from multinational regulators and various litigations), this is misleading. A close reading of Google's statements and policies indicates the company does not disclose the full extent of the information it collects on consumers, nor the valuable inferences it draws from this data. Analysis of communications from an Android smartphone suggests Google keeps hidden far broader profiles on billions of consumers around the world – removed from individual view or access, and public accountability. For example Google's "My Activity" page contains a history of what the consumer viewed, searched for, and browsed.¹ However, it *omits* much of the data the company collects, which is often far more invasive and revealing.

This *omitted* data is a consumer's "shadow profile" – massive, largely hidden datasets of online and offline activities. This information is collected through an extensive web of Google services, which is difficult, if not impossible to avoid. It is largely collected invisibly and without consumer consent. Processed by algorithms and artificial intelligence, this data reveals an intimate picture of a specific consumer's movements, socio-economic status, demographics, "likes," activities and more. It may or may not be associated with a specific users' name, but the specificity of this information defines the individual in such detail that a name is unnecessary.

Google offers a "Takeout"² page that purports to offer a complete view of the data Google collects on a consumer. Consumers can download a file including "Takeout data," which

¹ <https://myactivity.google.com>

² <https://takeout.google.com>

includes the content that Google scans to infer personality and interests, such as emails, interactions with other consumers, ad clicks, location, uploaded documents, and physical activity data. However, this file, which can contain years of personal information, omits *entire categories* of other data collected by Google.³ While purporting to provide a complete picture of the data Google holds on a consumer, it is only a fraction of Google’s actual online tracking.

Notably, “Takeout” data excludes a consumer’s interest profiles, the most critical information that Google stores. Google only shows users the interests that it ascribes based on their personal data *if the consumer elects to see personalized ads* from Google. Yet Google’s data set is so immense and its collection so pervasive that it can profile the interests of, and deliver ads to, consumers who have “opted-out” or deleted their data *just as effectively* as it can to consumers who remain inside Google’s ecosystem. This information is not included in “Takeout” data, leaving consumers in the dark.

Furthermore, the Takeout service only works for consumers who have a Google Account. Consumers who are *not signed into, or do not even have*, a Google Account may still have data collected on them and remain subject to Google’s privacy policy and terms of service. A consumer visiting a website using Google Analytics is automatically subject to Google’s privacy policy (data collection policies), allowing Google to collect unique identifiers on their device, their location, “cookie” data and metadata. None of this data is accessible or known to the consumer.

This data collection keeps Google in business. Google monetizes both the data available in “Takeout” and “shadow profile” data through digital advertising. For example, in communications to advertisers and publishers, Google highlights its ability to target ads based of Internet Protocol (IP) Address. Google also admitted that it *infers demographic data* from a consumer’s IP Address.⁴ Google tells advertisers it is able to tie this profile to consumers via cookies.⁵ As a result, in 2020, Google’s advertising revenue totaled \$147 billion, or 80%⁶ of its total. The more data Google collects on consumers, the better it can target ads and the more money it makes.

Google’s Android Data Collection Platform

One of the most *invasive and pervasive* tools in Google’s data collection arsenal is the Android smartphone. Smartphones are so integrated in consumers’ everyday life that it is literally an extension of a consumer’s personality. For sure, a smartphone is a phone, a calendar, a web browser, a music player, a camera and an access point for social media, but it is also an invasive tracker of health, precise movements, location, interests, and places frequented. Further,

³ Including people’s webpage interactions, ad interactions, device sensor data (eg: from their Android phones), search results clicked, Chromecast usage data, Google Docs keywords, Email keywords, and social graph.

⁴ Letter from Google to US Senate (Page 5, Paragraph 3)

<https://www.blumenthal.senate.gov/imo/media/doc/05.11.2018%20-%20FTC%20-%20Google%20Location%20History.pdf>

⁵ <https://support.google.com/google-ads/answer/2580383?co=ADWORDS.IsAWNCustomer%3Dfalse&hl=en>

⁶ <https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html>

smartphones are also surveillance tools for Google to collect important information about one’s physical environment, such as nearby Wi-Fi base station or Bluetooth beacons, in both public and private places. Google not only constantly tracks the location of Android users, but also links the data collected by an Android smartphone to unique device and account identifiers. (Table 1)

Unique Identifier	Description	Scope
Name and Email Address	Individual’s ID and user name in the Google ecosystem	Account
Advertising ID	ID for advertising, provided by Google Play Services	
Android Certificate	Signifies a Google account on a device is verified ⁷	
International Mobility Equipment Identity (IMEI)	Universal hardware identifier for mobile phone	Device
Media Access Control (MAC)	Hardware identifier for devices on a network	
Internet Protocol (IP) Address	Every device connected to the internet is assigned an IP address ⁸ ; can be used to establish a device’s location	
Serial Number	A manufacturer specific hardware identifier	

Table 1: Unique Identifiers Associated with Devices and Google Accounts

One particular Google service on Android smartphones – “checkin” – ties together many unique identifiers Google collects about a consumer and their device(s). With this data Google can readily combine multiple sets of data into a large super profile of a consumer. For example, Google’s Android tracks a mobile phone’s unique IMEI, linking it in the same file communicated to Google’s servers to account identifiers such as an Android ID, (Figure 1) which begs the question whether is it more important to know a consumer’s name or their unique set of IDs.

⁷ <https://developers.google.com/android/guides/client-auth>

⁸ <https://policies.google.com/privacy?hl=en>

conv ID	621864
test name	0107WeatherChannelTest
phone	Sony Xperia A
host	android.googleapis.com
api	checkin
src file	
timestamp	2019-01-16 14:00:05
— from req —	
imei	357008084578329
wifi_mac	84c7ea832819
android_id	3ff08958e43ca784
email/acct	[andrea.x.test@gmail.com]
— from req —	
model	G8142
device	G8142
serial	CB512F35YH
fingerprint	Sony/G8142/G8142:8.0.0/4...
manufacturer	Sony

Figure 1: Example Data Elements Reported by Google Android via “checkin” Service

Reviewing network communications between an Android phone and Google servers, at least four different types of identifiers are transmitted, collecting at least 18 different data elements. (Table 2) Google combines the data it collects about account and device identifiers with accurate and specific location information of a consumer. Location data linked with an Android ID and/or other unique identifiers including a consumer’s Google account is personally identifiable. Over time, this data creates a detailed profile about a consumer; where they live, work, shop, eat, socialize with, and many other revealing insights about their pattern of life, for Google’s use in providing detailed advertising profiles.

Type of Data	Data Elements Collected
Device Identifiers	Make, Model, Manufacturer, Android Version
Unique Device Identifiers	Serial Number, International Mobility Equipment Identity (IMEI), Media Access Control (MAC) Address, Internet Protocol (IP) Address
Google Account Identifiers	Email Address, Android ID, Advertising ID
Location and Sensor	GPS, Wi-Fi Access Points, Bluetooth Beacons, Activity Readings, Barometric Pressure, Gyroscope, Accelerometer, and Magnetometer Readings

Table 2: Key Data Types and Elements Collected by Google Android

Through constant tracking of consumers in the physical world and on the internet across various devices, Google is able to create a virtual dossier on nearly every internet user for the purposes of digital advertising and developing new products and services. The myriad of app level, device level, account level collection, combined with numerous *redundant* ID’s creates a cat and mouse game where consumers – even the most sophisticated consumers – reveal far more data than they

intend. Google’s vast data set on consumers is critical to its ability to generate revenue via advertising.

Data Missing from Google Takeout

Google claims consumers have control of their data via Google Takeout, a service available to Google Account holders to “create an archive with your data from Google products.”⁹ As stated above, the data Google makes available to consumers through this process is a limited portion of the larger super-profile that Google maintains on consumers. (Table 3) As evidenced by network transmission logs from Android devices, there are specific gaps between what a Google Android user’s device collects and the information Google reveals in a consumer’s Takeout data. Missing data includes information on nearby Wi-Fi base stations and Bluetooth beacons used to establish location, despite the fact this data is directly linked to a Google account at the time of collection. This missing information provides essential data for Google’s “shadow profile” on consumers.

Location Data Element	Collected by Google?	Tied to Unique Identifier?	Type of Identifier	Available to Takeout?
GPS Coordinates + Accuracy	YES	YES	Android ID	PARTIAL
Altitude	YES	YES	Android ID	PARTIAL
Activity Readings + Confidence Level	YES	YES	Android ID	YES
Wi-Fi Scans	YES	YES	Android ID	NO
• MAC Address	YES	YES	Android ID	NO
• Signal Strength + Frequency	YES	YES	Android ID	NO
Bluetooth Beacon Scans	YES	YES	Android ID	NO
• MAC Address	YES	YES	Android ID	NO
• Signal Strength + Frequency	YES	YES	Android ID	NO
Cell Tower Readings	YES	YES	Device Fingerprint ¹⁰	NO
Accelerometer Readings	YES	YES	Device Fingerprint	NO

⁹ <https://takeout.google.com/settings/takeout>

¹⁰ A device’s fingerprint is comprised of attributes, when combined, identify that device.

Magnetometer Readings	YES	YES	Device Fingerprint	NO
Gyroscope Readings	YES	YES	Device Fingerprint	NO
Barometric Pressure Readings	YES	YES	Device Fingerprint	NO
Source of Location Reading (Cell or Wi-Fi)	YES	YES	Android ID	NO
Connection to Wi-Fi Access Points	YES	YES	Android ID	NO
IP Address	YES	YES	Various¹¹	NO
PlaceIDs	YES	YES	Android Cert	NO
Rate + Change in Rate of Collection	YES	YES	Android ID	NO

Table 3: Data Missing from Google Takeout

Google’s Privacy Policy details how Google makes use of data collected from Wi-Fi Access Points, Bluetooth Beacons, and even a consumer’s IP Address to accurately locate a consumer.¹² To collect this data, *Google opts consumers into* extensive location tracking by default (Figure 2) when creating an account. Yet when an individual requests their data through the Google Takeout process, Google does not acknowledge or report the Wi-Fi, Bluetooth, and IP address data that Google collects.

¹¹ IP addresses are assigned by ISPs or Wireless Carriers to subscribers' devices. The address groups vary geographically, so an IP address can be used to estimate a device's location.

¹² <https://policies.google.com/privacy?hl=en#infocollect>

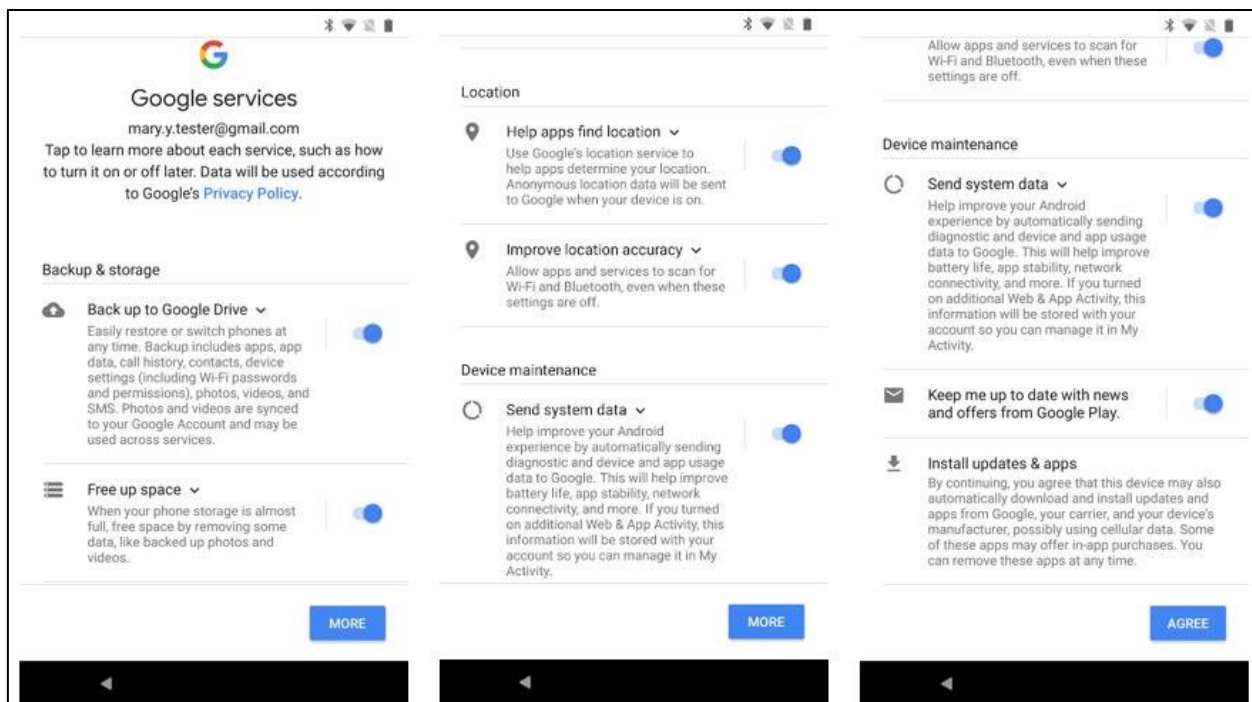


Figure 2: Screenshots of Google Services Defaults During Android Device Setup

For example, in the Location History file contained within the Takeout documents provided to a consumer, Google reports a list of GPS Coordinates and altitude readings accompanied by an accuracy range of that location in meters and timestamp. (Figure 3) While this information is some of the data Google collected about the device, it is not a comprehensive list of the location data and metadata associated with a consumer.

Figure 3 compares the information presented to the consumer from Google's Takeout documents, with a copy of the network communication to Google servers from the same Android device during this same time period. While Google collects, scans, and stores barometric pressure readings, Wi-Fi base stations and Bluetooth beacons via Android devices to determine the location of a consumer,¹³ it does not make that data available, even though the information is directly tied to a consumer's Google Account. Figure 3 reveals details on the location event recorded by an Android device via a Wi-Fi Scan, yet in the Takeout documents Google does not reveal the source of location or the list of Access Points used to pinpoint the location.

Location information is valuable to Google for the purposes of targeted advertising. Exact GPS coordinates are a very precise way to locate a consumer, but GPS is both taxing on the battery of a device and does not work indoors (for example, shopping malls). By scanning and collecting unique identifiers (in this case an Android ID) and the signal strength of Wi-Fi base stations near

¹³ Barometric pressure readings inform the altitude of the device and Wi-Fi scans inform the location reading, 38.877215, -76.9975140.

the device, Google can precisely calculate a consumer's location wherever they move in the world.

Creating an up-to-date map of Wi-Fi base stations globally sounds daunting, but with more than 2 billion¹⁴ active global Android users, Google can maintain a detailed database of access points updated constantly by the movements of unwitting consumers. Google's Bluetooth beacon database works in the same manner. And to ensure devices are located on the correct floor of a multi-story mall, Google uses the barometer data from Android devices to determine consumers' altitude. Clearly this data is valuable for Google, and it is collected directly from consumers' Android smartphones – however this data is missing from the Google Takeout documents.

Data about a consumer's movement and pattern of life allows Google to infer sensitive and unique information about consumers. Figure 4 is an example of a small amount of data collected by Google that initially seems benign (a record listing the Wi-Fi base station an Android device is connected to, along with a timestamp). Yet, if a consumer connects to the same Wi-Fi access point at 9 AM Monday-Friday, the Wi-Fi base station likely represents the consumer's place of work. Similarly, if a consumer connects to the same Wi-Fi base station every day at 7 PM and stays connected through the evening, the Wi-Fi base station is likely in located in the consumer's home.

```
{
  "timestampMs": 1550094845569,
  "wifiConnectivityStatus": {
    "mac": 123597800553519,
    "wifiConnectionStatus": "CONNECTED"
  }
}
```

Figure 4: Android Device reporting Wi-Fi connection to Google

Google also records when the data collection rate on an Android device changes – an indication when a consumer is using or moving with the device. Figure 5 highlights two types of rate change records – an average or “Normal” rate (left), and a “stationary” rate (right). Just as with Wi-Fi base stations, Google can infer useful information from this seemingly benign data collection. If an Android smartphone is set for a normal rate of data collection until 10 PM every day, but then switches to a stationary rate of data collection until 5 AM, the consumer is likely asleep between these hours. When combined with data on Wi-Fi base stations (Figure 4), patterns of life can be readily inferred. However, Google does not provide the data it collects about connections to Wi-Fi base stations or changes in data collection rate to a consumer via its “Takeout” service.

¹⁴ <https://www.theverge.com/2017/5/17/15654454/android-reaches-2-billion-monthly-active-users>

```

{
  "rate": {
    "description": "Normal",
    "timestampMs": 1549232208574,
    "uploadPeriodMs": 600000,
    "uploadReason": "default",
    "uploadSource": "internal"
  }
}

{
  "rate": {
    "description": "stationary",
    "newRequest": false,
    "samplePeriodMs": 2160000,
    "sampleReason": "stationary",
    "sampleSource": "internal",
    "timestampMs": 1549231848192
  }
}

```

Figure 5: Device reporting rate of data collection to Google

Google also collects timestamped data from the device’s location, orientation, and motion sensors. Figure 6 details the number of readings collected by Google from accelerometer, magnetometer, and gyroscope sensors at a rate of up to 100 readings per second. Data from these sensors can distinguish the types of activities a consumer is undertaking. For example, the device’s accelerometer can determine speed, distinguishing between activity types such as walking or riding in a car. Motion sensors can also indicate if a consumer is actively interacting with their Android device. For example, tracking the device’s orientation will indicate if a consumer positioned the phone in a manner where the screen is visible. None of this data is available to the consumer in Google Takeout.

sensor info - sensor info page +	
02-17 14:34:22 (EST)	baro 153 readings
02-17 14:34:19 (EST)	mag 1499 readings
02-17 14:34:19 (EST)	accel 1564 readings
02-17 14:34:52 (EST)	baro 158 readings
02-17 14:34:49 (EST)	mag 1505 readings
02-17 14:34:49 (EST)	accel 1556 readings
02-17 14:35:21 (EST)	baro 148 readings
02-17 14:34:19 (EST)	gyro 4164 readings
02-17 14:35:18 (EST)	mag 1508 readings
02-17 14:35:18 (EST)	accel 1553 readings
02-17 14:35:51 (EST)	baro 148 readings
02-17 14:34:49 (EST)	gyro 4149 readings
02-17 14:35:48 (EST)	mag 1501 readings
02-17 14:35:48 (EST)	accel 1544 readings
02-17 14:36:20 (EST)	baro 147 readings
02-17 14:35:18 (EST)	gyro 4139 readings

Figure 6: Android Device Reporting Motion and Orientation Sensor Data to Google

The fact is that notwithstanding Google “Takeout,” the information Google retains for itself is redundant such that it is as valuable for Google in targeting and tracking consumers for ads.

In or Out - Google Collects and Uses Data at All Times

According to Google’s privacy policy, “you can use many Google services when you’re signed out or without creating an account at all.”¹⁵ Importantly, Google collects data on consumers even if the consumer does not have an account or is signed out:

*when you’re not signed in to a Google Account, we store the information we collect with unique identifiers tied to the browser, application, or device you’re using.*¹⁶

If a consumer with a Google Account signs out of Google services or attempts to use a feature of a Google Chrome web browser known as “Incognito Mode” (a supposedly privacy protective browsing mode Google markets), Google still tracks the consumer. In a response to a question from the United States House of Representatives Judiciary Committee Google CEO Sundar Pichai explains: “When a user conducts a search on Google in Chrome Incognito and signed-out modes, we set a cookie to correlate searches conducted in the same Incognito window during the same browsing session.”¹⁷ Pichai continues, “We will, however, use certain factors” ... “such as the browser type, language, time of search, location (or an estimation of location), and prior browser session searches, to improve Search ranking relevance for the user’s query.”¹⁸ Google is still tracking the consumer via unique identifiers as outlined in their Privacy Policy, but never makes this data available to the consumer using Takeout.

How Google Collects and Uses Data on Consumers without a Google Account

Mr. Pichai’s explanation of how Google tracks consumers signed out of their Accounts or in incognito mode also provides insight into how Google tracks consumers who may not have a Google Account at all. Google still tracks these consumers when they interact with Google Services that do not require an account, such as Search or YouTube.

Google profits from a number of services and products that are not directly consumer facing and do not require a Google Account to use all of which are governed by Google’s Privacy Policy.¹⁹ For example, by shopping on Jcrew.com²⁰ or reading the news at NewYorkPost.com,²¹ a consumer consents – unknowingly – to the Google Privacy Policy as those websites use Google Analytics. It is difficult, if not impossible, to use the internet without encountering Google Analytics as approximately 75% of the top 100,000 websites on the internet use Google Analytics.²² In other words, consumers merely visiting websites on the

¹⁵ <https://policies.google.com/privacy?hl=en#infocollect>

¹⁶ <https://policies.google.com/privacy?hl=en#infocollect>

¹⁷ Google CEO Sundar Pichai’s response to US House Judiciary Questions for the Record. Page 3, Question 5.

¹⁸ Google CEO Sundar Pichai’s response to US House Judiciary Questions for the Record. Page 3, Question 6.

¹⁹ Google’s CEO Sundar Pichai’s testimony to the U.S. House Judiciary Committee <https://www.c-span.org/video/?455607-1/google-ceo-sundar-pichai-testifies-data-privacy-bias-concerns&start=11932> (3:18:52)

²⁰ https://www.jcrew.com/help/cookie_policy.jsp

²¹ <https://nypost.com/privacy/>

²² <https://trends.builtwith.com/analytics/Google-Analytics>

internet have a 75% chance of being captured by Google’s privacy (data collection) policy even if they have no other direct link to Google. Per its 2016 change in privacy policy, Google can then combine all of the data from its analytic properties with data generated by consumers using Google services to create a super-profile.²³

Google explains to advertisers how this process occurs on its “demographic targeting” help page for ads:

“For people who aren't signed in to their Google Account, we sometimes estimate their demographic information based on their activity from Google properties or the Display Network. For example, when people browse YouTube or sites on the Display Network, Google may store an identifier in their web browser, using a “cookie.” That browser may be associated with certain demographic categories, based on sites that were visited.”²⁴”

Google tracks a consumer across sessions and stores the data they generate. For example, one of the features for Google Developers is a function called PlaceIDs.²⁵ (Figure 6) Consumers using Google Maps see different places populating the map, depending on the demographic data Google has collected about them. Google plainly explains to developers how invasive and profound Google’s data collection is by remarkably stating on its Maps API documentation:

“Every visitor to your site sees a Google Map tailored just for them”²⁶

Regardless of a consumer’s Google account status or Location History setting, Google’s algorithms determine which places to show each consumer based on a super-profile informed by the data Google collects. For example, if a signed-out user opens Google Maps and searches for Breckenridge, Vail and finally Tahoe, the user is likely to see a specific ski resort populate the map.

This data is linkable to an individual via “unique identifiers” such as an Advertising ID or an IMEI²⁷ or the cookies described by Google above. This data can also be tied to the consumer’s location at the time of the search via their IP Address.²⁸

In addition to being linked to a consumer through various unique identifiers, highly specific location data is unique to an individual over time. Google affirms this conclusion by offering advertisers the ability to serve highly targeted digital ads based on consumers’ location, regardless of those consumers’ Google account status.

²³ https://www.google.com/intl/en_US/policies/privacy/archive/20160325-20160628/

²⁴ <https://support.google.com/google-ads/answer/2580383?co=ADWORDS.IsAWNCustomer%3Dfalse&hl=en>

²⁵ <https://developers.google.com/places/place-id>

²⁶ <https://developers.google.com/maps/documentation/embed/guide>

²⁷ <https://policies.google.com/privacy?hl=en#footnote-unique-id>

²⁸ <https://support.google.com/google-ads/answer/2453995?hl=en>

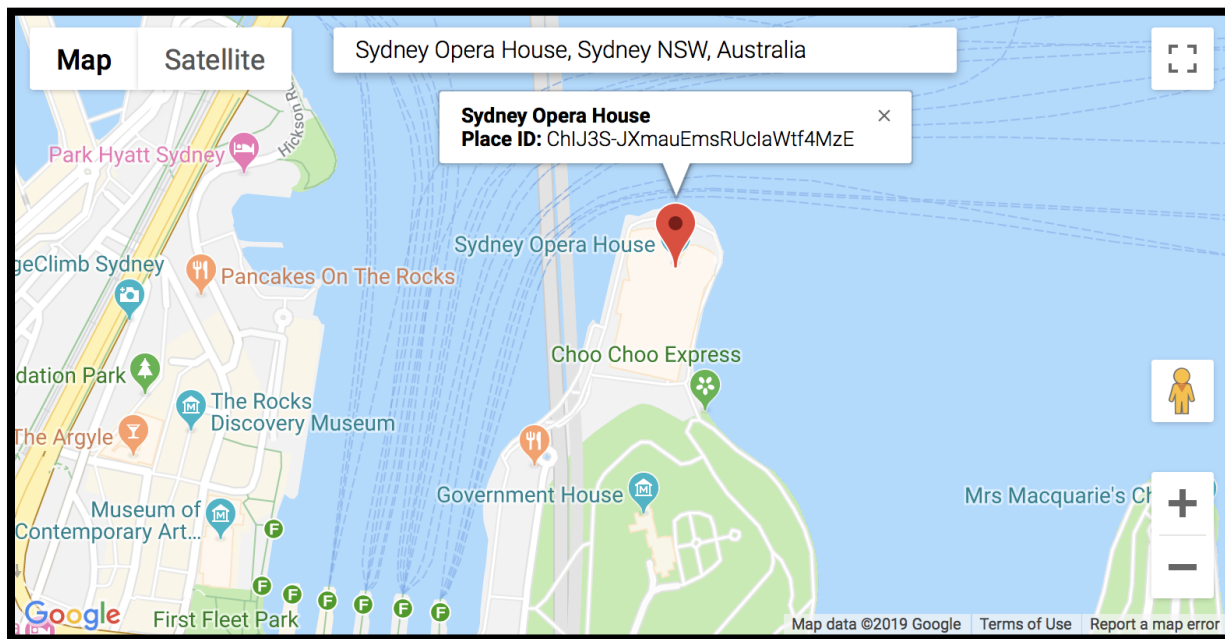


Figure 7: Google PlaceID for Sydney Opera House

Data Google Collects about the World (via Consumers)

When a mobile Android device sends Google a consumer's location, Google is able to maintain a self-updating and highly accurate map of devices moving throughout the world which can locate consumers in relation to various PlaceIDs on a map. Google claims to tell with 99% accuracy if, after seeing a digital advertisement for a store, a consumer enters the physical store location.²⁹ They can make this claim because Google has a detailed map of consumers' movements, data on the dimensions of millions of retail locations,³⁰ and a database of PlaceIDs.

Evidence of Google's constant location tracking are apparent in some of its consumer-facing products, such as Google reviews. (Figure 7) For example, the Google reviews of Sydney Opera House indicate the busiest times to visit are Friday and Saturday nights between 7 and 9 PM; this data is based off of visits to the location tracked surreptitiously on a consumer's smart phone, aggregated with the history of all consumers who have visited this location. Because Google has the world's most extensive database of places corresponding with specific locations, Google can link information about multiple devices at one location to assess busiest or most popular times for a given place.

²⁹ <https://static.googleusercontent.com/media/www.google.com/en/us/adwords/start/marketing-goals/pdf/white-paper-bridging-the-customer-journey.pdf>

³⁰ <https://adwords.googleblog.com/2016/09/New-Digital-Innovations-to-Close-the-Loop-for-Advertisers.html>

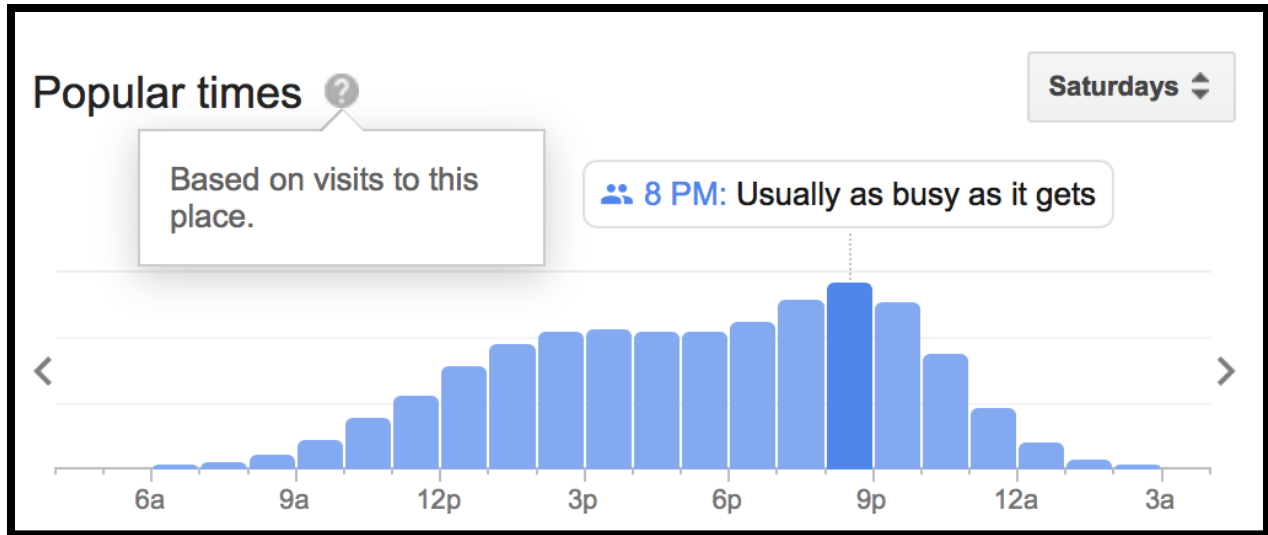


Figure 8: Google Reviews of Sydney Opera House

Google Promotes its Consumer Location Tracking Capabilities to Advertisers

Despite public claims that Google “builds privacy that works for everyone,”³¹ Google’s business model works to provide *everything but privacy* for the consumer. Google generates the majority of its revenue through advertising, powered by its ability to generate and combine large amounts of specific consumer data about consumer behavior on the internet with real time consumer activity and location data from mobile phones, as well as a myriad of other surreptitious collection points, such as internet cookies and application metadata. Google uses the location data it collects from mobile devices over time to establish patterns of life for consumers and acknowledges tracking both signed in and signed out consumers to infer interest in a location and inform a profile for the purposes of selling ads.

Google promotes its capability to target advertisements to a specific location in the world via IP Address and device location.³² (Figure 8) In order for Google to target advertisements by a user’s IP Address or the location of a device, it is reasonable to conclude Google must collect and tie this data back to individual consumers, feeding the result into digital advertising profiles.

Even though consumers are told they are “in control” of how Google collects and uses their data,³³ there is no way to disable location tracking via IP Address. Despite its relevance to consumers and advertisers, historical IP Address-based location information is not available to consumers via the Google Takeout service. Similarly, the catalogue of GPS and Cell Tower location data, Wi-Fi base station and Bluetooth beacon scans Google uses to locate consumers is not available via Takeout. And of course, none of this information is available to consumers who do not have a Google Account, but are uniquely identified and tracked by Google.

³¹ <https://safety.google/privacy/>

³² <https://support.google.com/google-ads/answer/2453995?hl=en>

³³ <https://safety.google/privacy/privacy-controls/>

Where your users are located (physical location)

The Google Ads system uses a number of factors to determine someone's physical location and whether to show your ad. When possible, we determine physical location based on someone's computer or mobile device location, or other methods.

- **IP address:**

Location is typically based on the Internet Protocol (IP) address, which is a unique number assigned by Internet Service Providers to each computer connected to the Internet.

If a device is connected to a Wi-Fi network, we may detect the Wi-Fi network's IP address to determine physical location. If a mobile device is connected to a mobile carrier's proxy server, we may use the carrier IP to determine the device's location.

- **Device location:**

Depending on a user's location settings, we may be able to use a precise location for advertising, based on one of these sources of location data:

- **GPS:** Accuracy varies depending on GPS signal and connection.
- **Wi-Fi:** Accuracy should be similar to the access range of a typical Wi-Fi router.
- **Bluetooth:** If Bluetooth and/or Bluetooth scanning are enabled on a device, a publicly broadcast Bluetooth signal can provide an accurate indication of location
- **Google's cell ID (cell tower) location database:** Used in the absence of Wi-Fi or GPS. Accuracy is dependent on how many cell towers are located within an area and available data, and some devices don't support cell ID location.

Figure 9: Google Ads Help Explanation Targeted Ads by Geolocation

Google also allows advertisers to target consumers based on their interest in a location, and highlights the value of tracking consumer location over time. (Figure 9).³⁴ Among other factors, Google infers a consumer's interest in a location based on their physical history at a particular location as well as searches on Google Maps.


³⁴ <https://support.google.com/google-ads/answer/2453995?hl=en>

Locations your users showed interest in (location of interest)



If the Google Ads system detects geographic areas that someone is interested in, we may show appropriate ads targeted to that area or surrounding areas (known as "location of interest").

Some of the ways that we might detect a location of interest are based on:

- Terms used in searches that indicate a location.
- Past searches that indicated a location of interest.
- A person's past physical locations.
- The content and context of a website where an ad is displayed. Keep in mind that the mention of a location on a page doesn't always indicate an interest in that location.
- Searches on Google Maps or Google Maps for Mobile.
- If someone [sets a custom location for Google search results](#) .

On the Google Display Network, we may infer a location associated with a page or site when we believe it will be useful for targeting your ads. A location mentioned on a page may not always indicate interest in that location. For example, someone who is reading news about San Francisco isn't necessarily interested in ads for San Francisco florists. Similarly, we might infer an interest in a location, even if that location isn't specifically mentioned on a page, but the context of the entire site indicates an interest in that place.

Location of interest isn't restricted to a person's country, or the Google search domain the person is searching on. For example, if someone in Paris, France searches for *Los Angeles taxi* on [google.fr](#) (France), we'll still identify Los Angeles as a location she's interested in.

Figure 10: Google Ads Help Explanation Targeted Ads by Geolocation

From Data to Dollars

A consumer's pattern of life – the daily rhythm of the people and places individuals spend time in the real world – combined with online web browsing, search history and a myriad of other data points creates an intimate dossier of a consumer's lifestyle. Google uses this data to develop and continuously update its super-profile on consumers. Combining multiple sources of user data across its products, services devices and accounts, the pool of data is used to power Google's digital advertising, responsible for 86% of Google's revenue.³⁵

Through its various digital services, Google is able to track consumers across the internet. These services include what a consumer can directly link to their Google Account, (Search, YouTube, Gmail, Hangouts, etc.) as well as various AdTech and Analytic Products where a consumer may not have a direct relationship (Google Ads, DoubleClick, Google Analytics, etc). Google does not claim to target individual consumers with specific ads, rather, Google works at a larger scale - targeting *all* consumers based off of their individual demographics, location, and intent. Within

³⁵ https://abc.xyz/investor/static/pdf/20171231_alphabet_10K.pdf?cache=7ac82f7

Google’s ad products, there are multiple ways to target specific advertising “audiences.” According to Google, audiences are “groups of people with specific interests, intents, and demographics, *as estimated by Google*” (emphasis added).³⁶ Most broadly, an advertiser can target an ad campaign based on various demographic data points.³⁷ In a letter to the US Senate, Google explains how it infers demographic data based on a consumer’s location for the purposes of advertising.³⁸

Audience specific targeting allows advertisers to reach consumers based on their individual interests, a feature Google calls “affinity audiences”³⁹ (Figure 10), as well as their intent, called “in-market audiences”.⁴⁰ (Figure 11) Google makes assessments of a consumer’s affinity and / or intent based off of the data collected on consumers via their interaction with Google’s Services (such as Android or websites that use Google’s advertising or analytic products). To further refine an audience, an advertiser can target websites related to varying subjects, called “targeted topics”.⁴¹ (Figure 12)

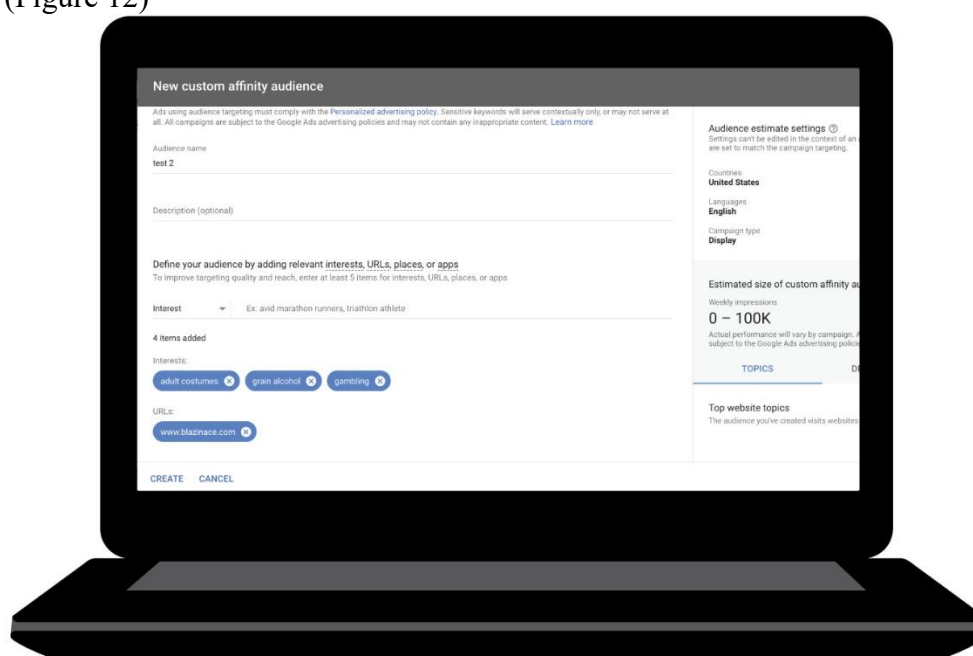


Figure 11: Sample Affinity Audiences on Google Ads Platform

³⁶ <https://support.google.com/google-ads/answer/2497941?hl=en>

³⁷ <https://support.google.com/google-ads/answer/2580383?co=ADWORDS.IsAWNCustomer%3Dfalse&hl=en&oco=0>

³⁸ Letter from Google to US Senate (Page 5, Paragraph 3) <https://www.blumenthal.senate.gov/imo/media/doc/05.11.2018%20-%20FTC%20-%20Google%20Location%20History.pdf>

³⁹ https://developers.google.com/adwords/api/docs/appendix/affinity_categories.csv

⁴⁰ https://developers.google.com/adwords/api/docs/appendix/in-market_categories.csv

⁴¹ <https://support.google.com/google-ads/answer/2497832?hl=en>

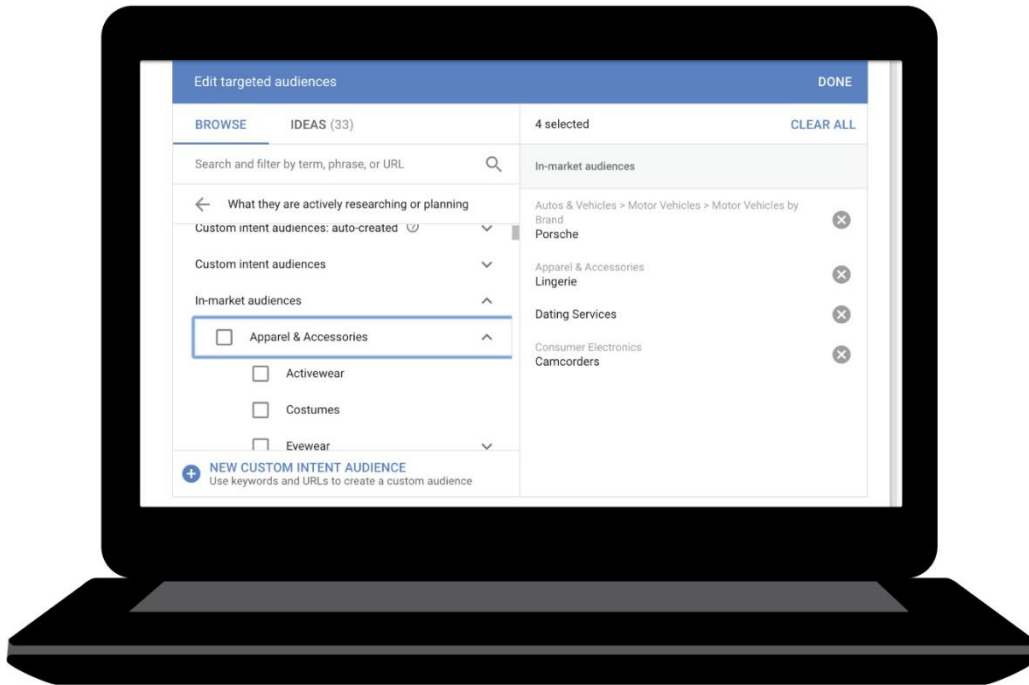


Figure 12: Sample In-Market Audiences on Google Ads Platform

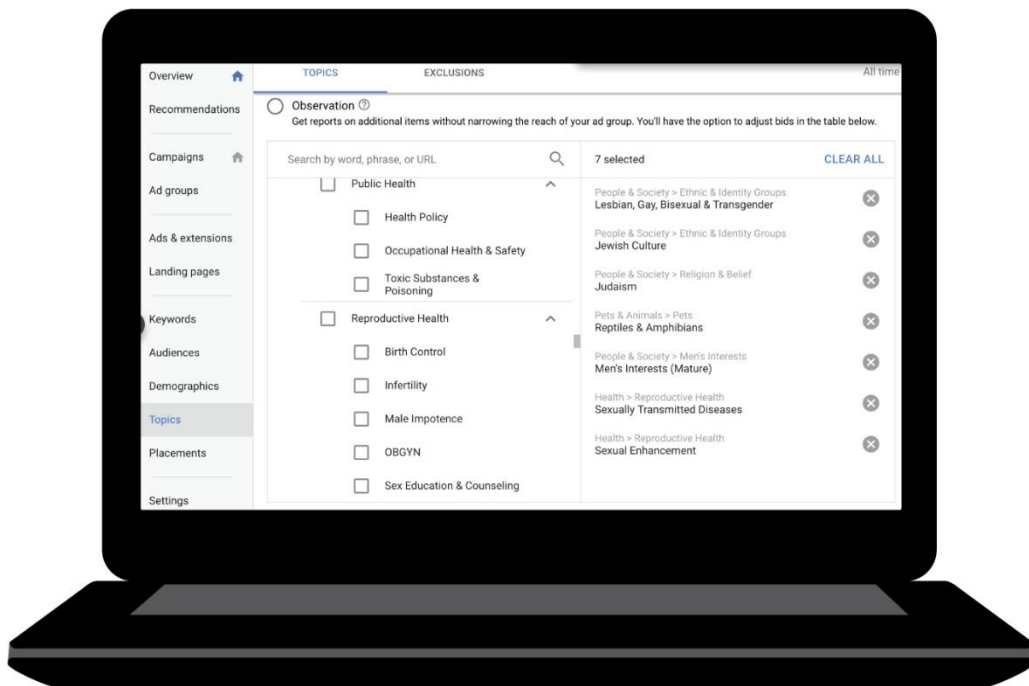


Figure 13: Sample Targeted Topics of Google Ads Platform

Advertisers have the ability to combine data sets across demographic, affinity, and intent audiences. For instance, an advertiser could target a 40-year-old married male with children that makes \$50k a year whose interests are:

- “adult costumes”
- “grain alcohol”
- “gambling”

The same advertiser could then combine this profile with an intent such as:

- “gentlemen’s club”
- “infectious diseases”
- “male enhancement”

To build upon the above example, an advertiser could combine the outlined profile of an individual with topics that same consumer might interact with on the internet. Google’s ad product provides options for topics such as:

- “divorce & separation”
- “depression”
- “male impotence”

By targeting against websites with specific topics, an advertiser can also indirectly target an audience. Any consumer’s interaction with a “topic” would also feed their “audience” profile. If the hypothetical consumer described above interacts with advertisements Google served against their interests, intent, and online activity, the interaction itself will add to the profile of this consumer, data that Google can use to improve targeting for advertisers.

Once a consumer interacts with an advertisement, the advertiser can add that individual to a “remarketing list,” allowing direct targeting of the individual in subsequent ad campaigns. Google also offers the ability to target “similar audiences” “to people who share characteristics with people on your existing remarketing lists”.⁴² Google’s data collection can reveal individuals who just *may* be interested in a product or service simply because that consumer is *similar* to another consumer who has demonstrated interest in a product.

Summary

Google’s business is designed to collect as much data as possible about as many consumers as possible. Yet only a small amount of the data Google collects is made available to Google account holders even though Google claims to provide an exhaustive list of the data collected. In reality, Google collects and stores significantly more data about each consumer, if they have a Google account or not, and ties this data to unique identifiers which enable it to link information back to an individual. The information that Google does not reveal to a consumer is that consumer’s shadow profile. As a result, consumers do not fully understand all of the data Google holds, which Google uses to target consumers at any moment across various products and services and is continuously updated as consumers navigate the internet and the real world.

⁴² <https://support.google.com/google-ads/answer/7139569?co=ADWORDS.IsAWNCustomer%3Dtrue&hl=en&oco=0>