Understanding the Scope of Data Collection by Major Technology Platforms

May 2020

Justin Brookman

CR Consumer Reports | Digital Lab

Investments from our members and philanthropic organizations including the Alfred P. Sloan Foundation, the Ford Foundation, and Craig Newmark Philanthropies support CR's efforts to promote consumer interests in relation to privacy, security, and data practices. This report was made possible by a grant from the Alfred P. Sloan Foundation.

I. Introduction

In 2018, in partnership with the Alfred P. Sloan Foundation,[1] Consumer Reports[2] initiated a project to develop a comprehensive understanding of the scope of data collection of the biggest technology companies. CR recorded and analyzed the data practices (collection, use, sharing, and control) of the organizations that organize our digital lives, the complex relationships they have with one another, and the level of insight consumers have to make informed choices.

The internet was built on principles of trust, decentralization, and anonymity. As the consumer internet has grown, these principles have been buried underneath new layers of complexity in support of products and services that consumers make use of for business, pleasure, and creativity. The data brokerage and advertising infrastructure that has become the economic engine of the consumer internet cannot be divorced from the daily conveniences users have come to expect. In 2014, early Internet entrepreneur and MIT academic Ethan Zuckerman wrote: "I have come to believe that advertising is the original sin of the web. The fallen state of our Internet is a direct, if unintentional, consequence of choosing advertising as the default model to support online content and services."[3] Though the internet makes our lives easier and gives us experiences we enjoy, consumers increasingly lack (i) insight into how decisions are made on our behalf and the subtle ways the marketplace can be manipulated, and (ii) robust public institutions that can keep up with the most fast-changing sector in the economy.

Yet, in response to an ocean of privacy scandals and data breaches, and a tech economy fueled by opaque data sharing, consumers are showing greater concern for their privacy than ever before.

Consumers need ways to examine and control how their data is collected, analyzed, and used by the organizations that provide the products and services they bring into their lives. Consumer Reports believes the first step toward enabling this control and a more transparent digital economy—and empowering consumers' choices—is to define

[1] Consumer Reports's donors and philanthropic partners play a critical role in our efforts to promote consumer interests in relation to privacy, security, and data practices. Our work on these issues is made possible by the vision and support of the Ford Foundation, the Alfred P. Sloan Foundation, and Craig Newmark Philanthropies. Craig Newmark is a former board member of Consumer Reports.

[2] Consumer Reports is the world's largest independent product-testing organization. It conducts its advocacy work in the areas of privacy, telecommunications, financial services, food and product safety, and health care, among other areas. Using its dozens of labs, auto test center, and survey research department, the nonprofit organization rates thousands of products and services annually. Founded in 1936, Consumer Reports has over 6 million members and publishes its magazine, website, and other publications.

[3] Ethan Zuckerman, *The Internet's Original Sin*, The Atlantic (Aug. 14, 2014), https://www.theatlantic.com/technology/archive/2014/08/advertising-is-the-internets-original-sin/376041/.

and understand the scope of data collection currently underway. This is the basis for our work with the Sloan Foundation.

The internet looks different for each consumer depending on where they live, who provides their service, and what technologies they can access. Technology platforms use consumer data to provide richer experiences for their users; it is helpful for a weather app to know your location or to receive news tailored to your interests. The challenge lies in how these choices are made for consumers without their knowledge or consent. Broadly speaking, consumers are concerned about their privacy and security online, but do not necessarily have the tools or understanding to exercise control. Consumers are complacent and confused about grappling with a system that provides them tremendous value, but requires a constant supply of personal data to operate—most of which is hidden or obfuscated to avoid scrutiny and public backlash.

II. Context and Methodology

Major technology platforms have unprecedented insight into our everyday lives due to our increasing reliance on connected devices, internet-based services, and cloud computing. Much of this data collection is relatively intuitive, such as online photo backups (e.g., Google Photos, iCloud Photos) and file storage (e.g., Dropbox, Google Drive). In other cases, as we use internet services to communicate with each other, companies are persistently collecting, storing, organizing, aggregating, transforming, and sharing our data, as well as the metadata used to facilitate that communication. Even when consumers intend to share data for a conscious purpose, it is not always clear for what secondary purposes the data is being collected and used, or with whom it is being shared.

Many data collection practices are less obvious, even if consumers are aware they are using a particular company's products. Consumers understand internet service providers transmit their internet traffic, but may not know what those providers track about their behavior. Operating systems and hardware have the technical capacity to observe and report back how consumers use devices and the software (systems, UI, etc.) on them. In other cases, consumers may have no indication that a company has visibility into their behavior; many web platforms provide unbranded functionality for websites and apps and can collect near comprehensive data about users' online behaviors.

Due to increasing vertical integration of services, major platforms have sophisticated capabilities to combine different aspects of consumers' lives into a composite picture. Companies can supplement this data with information purchased from third-party data brokers, with consumers being none the wiser. As such, it is difficult to describe this

practice as merely data collection, but as inference generation and profile-building. Unfortunately, the relationships between companies, their business partners, and their constituent organizations are not disclosed with enough detail to isolate and understand how an individual consumer is affected.

Scrutinizing any particular facet of these organizations is both difficult and incomplete without considering the entirety of their corporate structure and product portfolio, and their interconnectedness through technological integration and business relationships. This necessitates an approach that examines these organizations as collective *platforms* involved in facilitating the collection, transaction, and use of consumer data through products and services—and in many cases—advertising. This clarifies the research challenge but dramatically increases the scope. Additionally, this kind of research quickly becomes a fractal; more understanding leads to more questions. Therefore our examination concentrates on only a handful of the largest organizations deserving of scrutiny, but leaves out important components of the digital economy ecosystem, such as advertising and data brokers, banks, entertainment companies, etc.

In the end, this report analyzes the data behaviors of fifteen internet platforms:

- Alibaba
- Amazon
- Apple
- AT&T
- Charter
- Comcast
- Cox
- Facebook
- Google
- Microsoft
- Sprint
- T-Mobile
- Twitter
- Verizon
- Wikimedia

To develop a baseline understanding of these companies' data practices, we combined a technical review of data collected through smartphone applications with a detailed review of public documentation about internet companies' data behaviors. Consumer Reports and other technologists, academics, and experts are working to build the tools and methodologies required to peer deeper into the operation of these platforms; this is a central priority of our ongoing work. However, at the moment, our analytical capacity is limited to (i) desk review of policies and disclosures by the companies, (ii) desk review of secondary sources and research conducted by other experts, and (iii) technical assessment where the tools and access currently exist.

Technical Assessment: We partnered with AppCensus to measure the collection and sharing practices of personal information on 426 Android apps offered by these fifteen companies and their subsidiaries. This represents only a starting point for understanding the kinds of data harvested by these companies and their affiliates, and

the lengths some of them go to hide their activity.[4] This assessment focused not on the content of user interactions but on how the applications collected and shared other identifying information supplied by the phone. The study's findings, many of which may reveal violations of Android's developer terms, include:

- More apps collected and shared the Android ID than the more easily resettable Advertising ID;
- 40 apps collected hardware identifiers such as IMEI or IMSI[5];
- 24 apps collect the WiFi router's MAC address, potentially as a workaround for accessing geolocation without triggering a permission request;
- 24 apps transmitted the device's phone number; and
- 19 apps transmitted email addresses collected from the device.

Policy Review: For our policy review, we gathered nearly 140 primary sources (such as privacy policies, license agreements, FAQs, and privacy controls and dashboards), and over 120 secondary sources (such as news articles and academic papers) to determine what was publicly documented about these companies' privacy and data practices. However, this analysis is only a starting point supported by a representative selection of these documents that provide the most accessible and relatable information consistent across platforms: their privacy policies.

The data we collected demonstrates that most of these companies engage in far-reaching data practices, combining data sets from a wide array of disparate sources. Most of the companies we looked at did not describe affirmative limitations on their data collection and use, but instead provided non-exhaustive lists of the data (or *types* of data) they collect. Nevertheless, even these examples were overwhelming in many cases. Google alone (not to mention its parent company Alphabet) is comprised of at least 197 products and services (including hardware, web apps, development tools, and defunct services), each of which may have one or more documents describing their data practices. A single document we reviewed listed over 60 different data elements collected by Google, some of which can be quite extensive and illuminating (and opaque) all on their own, such as Search History or Browsing History[6]:

[4] The Android platform represents a relatively reliable and easily testable environment to establish our test suite for connected products and services. Over time, as part of our Digital Lab initiative, we will be expanding our dataflow measurement capabilities—for example, to include iOS and embedded devices—though certain practices that occur on remote cloud servers or directly between two companies will invariably be difficult to measure

[5] These hardware identifiers are difficult if not impossible for users to change as opposed to software-based identifiers like advertising identifiers.

[6] Privacy Policy: Information Google Collects, Google,https://policies.google.com/privacy?hl=en#infocollect.

Data Elements Collected by Google (illustrative)

"Hotword"
Ads clicked
Ads viewed
App information
Application version number
Birthday
Browser type and settings
Browsing history
Calendar events
Calling-party number
Comments on videos
Contacts
Crash reports
Date and time of search
Date of calls and messages
Date of request
Device type and settings
Duration of calls
Email Address
Emails received
Emails written
Forwarding numbers
Gender
Geolocation
Google account ID
GSP
Installed apps
IP address
Locations history
Metadata user search history
Mobile carrier name
Mobile network connection
Mobile number
Most recent sign-in time
Name
Operating system
Password
Payment Information
Phone Number
Publicly accessible information
Purchase activity
Recent searches
Referrer URL
Routing information
Safe search preferences
Search history
Search terms
Server Logs
System activity
Third party activity
Third party visits
Time of calls and messages
Time of requests
Types of calls
Unique advertising ID
Unique identifiers
URL of third party website
User cookies information
User network
User preferences
User-generated content
Videos watched
Voice and audio information
Voice query - content
Voice query - country
Voice query - language
WIFI network

Companies were also vague in most cases about the uses of this data, often reserving broad rights to use data for "research," "product improvement," or "advertising."[7] The companies did not provide many details about the behavioral categories or inferences made about users; although many platforms have that allow them to monitor the conclusions being drawn about their interests, some of these tools have disappeared or become less detailed over time. Our survey revealed that companies would sometimes

[7] *See*, *e.g.*, Change History for Microsoft Privacy Statement, Microsoft, https://privacy.microsoft.com/en-us/updates.

forbear from particularly egregious uses of data or from sharing data with third parties, but there were otherwise few affirmative constraints on what they themselves could do with user data. Thus, in many cases, our analysis merely identifies what these companies *could* be doing with data they collect, as U.S. law imposes few affirmative transparency obligations on companies to be forthcoming about all their data practices.

It is not clear from disclosure alone exactly what data is gathered or inferred, and then used for other purposes. The norm of describing data practices through example and vague description prevented us from being able to draw explicit relationships between data types and use of that data. For example, Amazon at the time of our review collected roughly 76 discernable pieces of information from users using various methods (via user, third parties, device, platform, user behavior, etc.), though only for 33 discernable categories were there descriptions of how they were used.

Amazon's Interest-Based Ads document describes Amazon's advertising technology and methodology, and includes a prime example of the ambiguities around collection and usage:

> Also, some third-parties may provide us information about you (such as the sites where you have been shown ads or demographic information) from offline and online sources that we may use to provide you more relevant and useful advertising.[8]

Presumably, since this statement comes from the Interest-Based Ads section of Amazon's website, we can assume that data, such as demographics, are given to Amazon by third parties for advertising. However, there is no clear statement to suggest that this data cannot be used elsewhere within Amazon.

III. A Day in the Life of Your Data

This paper organizes platform data collection by user behavior, describing how many different kinds of services are able to observe what we do in our everyday lives. The major categories of behavior that we investigated are: web browsing, mobile app usage, geolocation and other phone sensors, television viewing, internet of things, and offline shopping and other behavior. We conclude this section with a discussion of the methods that companies use to tie all of these behaviors together.

[8] *Interest-Based Ads*, Amazon Help & Customer Service, https://www.amazon.com/gp/help/customer/display.html?nodeId=202075050.

A. The Web Ecosystem

Consumers spend more and more of their lives online: the average U.S. adult spends nearly four hours a day looking at a computer, tablet, or smartphone—and many spend far longer.[9] More business and personal endeavors are conducted via computers and the internet than ever before. Despite the growth of other platforms, the World Wide Web continues to be a primary mechanism to engage with the outside world, and is an important component of the underlying technology that drives the applications we rely on for email, search, chat, banking, and more. And large internet platforms have a variety of ways of monitoring what consumers do in web browsers.

1. Websites

First, the companies that offer websites can collect a lot of information directly from users. In many cases, users input information directly through web forms; in other cases, the site's operator may simply passively observe how users engage with their service. This type of first-party tracking is relatively intuitive and consumers are more likely to have a contextual understanding of the direct data collection. On the other hand, consumers may not appreciate all first-party tracking, such as the use of JavaScript to monitor mouse movements or text entered into forms prior to submission.

Large internet platforms offer a wide variety of websites, many of which collect directly identifying information such as real name or email address,[10] along with an array of other sensitive personal data. Just a few examples include:

- Shopping and purchase behavior;
- Email and social media content;
- Credit card data;
- Home and business addresses;
- Search terms; and
- Health information.

Much of this data collection is functionally necessary to provide the services requested by an individual. However, it may be repurposed for other uses without transparency or

[9] Quentin Fottrell, *People Spend Most of Their Working Hours Staring at Screens*, Marketwatch (Aug. 4, 2018, 5:09 PM), https://www.marketwatch.com/story/people-are-spending-most-of-their-waking-hours-staring-at-screens-2018-08-01 [hereinafter *People Spend Most of their Working Hours*]; and, Andrew Perrin & Madhu Kumar, *About Three-in-Ten U.S. Adults Say They Are 'Almost Constantly' Online*, Pew Research Ctr. (July 25, 2019), https://www.pewresearch.org/fact-tank/2019/07/25/americans-going-online-almost-constantly/.

[10] Even data collected for security purposes for two-factor authentication may be repurposed for ad targeting. *See*, *e.g.*, Alfred Ng, *Facebook's Two-Factor Authentications Puts Security and Privacy at Odds*, CNet (Mar. 5, 2019), https://www.cnet.com/news/facebooks-two-factor-authentication-with-phone-numbers-puts-security-and-privacy-at-odds.

consent from the end user. Even if the user deletes data, information inferred from that data may be retained and used without the consumer's awareness.[11]

2. Ad Tech Tracking Companies

In addition to websites collecting information about you directly, most sites also share data with a number of third-party companies for a multitude of reasons, including advertising, optimization, and measurement. Modern websites embed code from dozens of third-party companies, in many cases enabling those companies to monitor your behavior across other sites that also embed their code.[12] These third-party companies typically use cookies[13] to track users across different sites: each time a new site sends information to a tracking company, the company can also see the unique cookie identifier it has stored on the user's browser, allowing it to associate the site visit with a particular user or device. Users have the ability to delete their cookies or block third parties from setting them in the first place; however, companies may use other techniques—such as digital fingerprinting—to track users without cookies.[14] Sharing with ad tech providers may also happen directly between companies' servers as well; in these cases, the data transfers would be undetectable to an end user.[15]

Many of the large platforms provide a range of services that other companies can embed into their sites, enabling the platforms to track users across those other sites. In recent studies, Google had a third-party presence on the vast majority of internet websites; Facebook had nearly as much penetration.[16] In recent years, some internet

[11] Todd Haselton, *Google Still Keeps a List of Everything You Ever Bought Using Gmail, Even if You Delete All of Your Emails*, CNBC (July 5, 2019, 1:12 PM), https://www.cnbc.com/2019/07/05/google-gmail-purchase-history-cant-be-deleted.html [hereinafter *Google Still Keeps a List*].

[12] *Meet the Trackers*, Me and My Shadow (Sept. 8, 2015), https://myshadow.org/trackography-meet-the-trackers; *and*, *see* Sebastian Shelter & Jéroôme Kunegis, *Tracking the Trackers: A Large-Scale Analysis of Embedded Web Trackers*, Proceedings of the 10th Int'l AAAI Conf. on Web & Social Media (2016), https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/download/13024/12827.

[13] On the other hand, companies may use other technologies to track web users as well, through similar cookie-like objects or digital fingerprinting. See *Consumer Information: Online Tracking*, Fed. Trade Comm'n (June 2016), https://www.consumer.ftc.gov/articles/0042-online-tracking; *and*, *see*, Dan Goodin, *Now Sites Can Fingerprint You Online Even if You Use Multiple Browsers*, Ars Technica (Feb. 13, 2017), https://arstechnica.com/information-technology/2017/02/now-sites-can-fingerprint-you-online-even-when-you-use-multiple-browsers/.

[14] Steven Englehardt & Arvind Narayanan, *Online Tracking: A 1-Million-Site Measurement and Analysis*, ACM CCS (2016), https://webtransparency.cs.princeton.edu/webcensus/ [hereinafter *1-Million-Site*].

[15] Emily Breuninger, *Pixels vs. Postbacks: Which Tracking Method Should You Be Using*, Tune (May 26, 2016), https://www.tune.com/blog/hasoffers-pixels-vs-postbacks-tracking-methods/.

[16] Justin Brookman et al., *Cross-Device Tracking: Measurement and Disclosures*, Proceedings on Privacy Enhancing Tech. (2017), p.133–148, *available at* https://petsymposium.org/2017/papers/issue2/paper29-2017-2-source.pdf [hereinafter *Cross-Device Tracking*]; *1-Million-Site*, *supra* note 14; Ibrahim Altaweel, Nathaniel Good, & Chris Jay Hoofnagle, *Web Privacy Census*, Technology Science (Dec. 15, 2015), https://techscience.org/a/2015121502/; Russell Brandom, *Google and Facebook Still Dominate Tracking*

service providers (ISPs) have acquired ad tech companies in order to compete with Google's and Facebook's ad businesses.[17]

Typically code from these companies' is embedded in order to facilitate advertising, but it may also serve other purposes such as analytics, measurement, or social sharing widgets ("like" buttons). Historically, most cross-site web tracking was *pseudonymous* in that cross-site browsing history was associated with a unique identifier but not a discreet real-name identity. However, because large platforms like Google and Facebook also collect identity information as a first party when users sign up and login to their services, they are able to associate third-party tracking data with particular individuals.[18] *See infra*, How Companies Tie All this Data Together. As such, these companies are able to generate an incredibly comprehensive view of what specific individuals do across websites and apps on different devices.

3. Browsers, Operating Systems, and Hardware

Web browsers have the capacity to monitor everything a user does on the web because they process the actual web addresses (URLs), renders them for view, and then acts as the agent for the user interaction with content. Most browsers allow users to browse in a "signed in" capacity, meaning that browsing history is synced across multiple browsers where a user logs in, and can also be associated with offline identity if linked to an email address or other identifier. Even when not logged in, most browsers report back some data to protect users from potentially malicious sites, or—usually after opting in—to provide the manufacturer with performance and crash data. In many cases, this can be done without transmitting or retaining data that links a particular user to specific content, but in some cases, browsers have been accused of transmitting back full url data along with device identifiers.[19]

Similarly, desktop and mobile operating systems also possess the technical capacity to observe and collect web browsing data. OS manufacturers' privacy policies are not

---

on the Web, The Verge (May 18, 2016), https://www.theverge.com/2016/5/18/11692228/google-facebook-web-tracking-survey-advertising.

[17] Ingrid Lunden, *Verizon Completes Its Acquisition of AOL for $4.4B*, TechCrunch (June 23, 2015), https://techcrunch.com/2015/06/23/verizon-completes-its-acquisition-of-aol-for-4-4b/; Alex Perry, *AT&T Acquires AppNexus: What the Merger Means for Consumers*, Int'l Bus. Times (June 25, 2018), https://www.ibtimes.com/att-acquires-appnexus-what-merger-means-consumers-2694577; Nat Levy, *Scoop: T-Mobile Acquires Mobile Marketing Startup PushSpring to Bolster Advertising Tech*, GeekWire (July 31, 2019), https://www.geekwire.com/2019/scoop-t-mobile-acquires-mobile-marketing-startup-pushspring-bolster-advertising-tech/.

[18] Julia Angwin, *Google has Quietly Dropped Ban on Personally Identifiable Web Tracking*, ProPublica (Oct. 21, 2016), https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking[hereinafter *Google Dropped Ban*].

[19] Script Junkie (@scriptjunkie1), Twitter (July 19, 2019, 2:14 PM), https://twitter.com/scriptjunkie1/status/1152280517972299777.

clear as to whether this data is collected (though they are typically broad enough to allow for such collection). Operating systems do increasingly rely on cloud processing and engage in collection of telemetry data to monitor performance, but there is no clear evidence that their collection is this detailed.[20]

Hardware manufacturers (even those who do not provide a device's operating system) may also have the capacity to monitor what a consumer does on the device: these manufacturers have the capacity to preinstall software on the device that may report back information about consumer behavior. It is not clear how widespread this practice is, but both computer[21] and phone[22] makers have at times been caught installing such software, including software capable of intercepting encrypted communications to external websites.

4. Internet Service Providers

Consumers' internet service providers have some ability to monitor customers' web traffic that flows over their networks, though they may not be able to view the full contents of traffic that is encrypted end-to-end.[23] Over 70 percent of internet traffic is now encrypted, which represents a massive expansion in the use of encryption in recent years.[24] Of course, even if web traffic is encrypted, ISPs may be able to discern a lot of information about a user's browsing. At the very least, an ISP will typically be able to monitor the web *domains* that a user is visiting (e.g., ESPN.com), even if they are not able to discern the particular url (e.g., https://www.espn.com/soccer/soccer-transfers/story/3908862/arsenal-beat-psgnapoli-to-sign-pepe-from-lille).[25] From this example, an ISP might be able to guess from the fact of visiting ESPN that a user was generally a sports fan, but not the more detailed knowledge that they were interested in soccer or the club Arsenal.

---

[20] Ed Bott, *Windows 10 Telemetry Secrets: Where, When, and Why Microsoft Collects Your Data*, ZDNet (Feb. 23, 2016), https://www.zdnet.com/article/windows-10-telemetry-secrets/.
[21] Russell Brandom, *Lenovo pays $3.5 Million for Preinstalling Superfish Adware*, The Verge (Sept. 6, 2017, 11:39 AM), https://www.theverge.com/2017/9/6/16261988/lenovo-adware-superfish-settlement-fine-state-ag; *Lenovo Settles FTC Charges it Harmed Consumers with Preinstalled Software on its Laptops that Compromised Online Security*, Fed. Trade Comm'n (Sept. 5, 2017), https://www.ftc.gov/news-events/press-releases/2017/09/lenovo-settles-ftc-charges-it-harmed-consumers-preinstalled.
[22] Zachary Lutz, *Carrier IQ: What It Is, What It Isn't, and What You Need to Know*, Engadget (Dec. 1, 2011), https://www.engadget.com/2011/12/01/carrier-iq-what-it-is-what-it-isnt-and-what-you-need-to/.
[23] Encryption is a "a digital tool that scrambles messages and other data to ensure that they cannot be accessed by anyone for whom they are not intended." *Beyond Secrets: The Consumer Stake in the Encryption Debate*, Consumer Reports (Dec. 21, 2017), https://advocacy.consumerreports.org/wp-content/uploads/2017/12/Beyond-Secrets-12.21.17-FINAL.pdf
[24] John Maddison, *Encrypted Traffic Reaches a New Threshold*, Network Computing (Nov. 28, 2018), https://www.networkcomputing.com/network-security/encrypted-traffic-reaches-new-threshold.
[25] *What ISPs Can See*, Upturn (Mar. 2016), https://www.teamupturn.com/reports/2016/what-isps-can-see.

Historically, ISPs did not use data derived from consumers' web traffic for much beyond security, safety, and fraud prevention.[26] However, in recent years, ISPs have increasingly offered analytics and advertising products based on monitoring web traffic.[27] Moreover, several ISPs have purchased online ad tech companies, seeking to use the data generated as an ISP to serve targeted advertisements on the web.[28] ISPs provide very little information in their policies about how long customer data is retained for various purposes; in March of 2019, the Federal Trade Commission announced it was conducting a study under Section 6(b) of the FTC Act to analyze the collection, use, retention, and disclosure practices of broadband service providers: that study may lead to greater transparency into ISPs' data practices.[29]

In 2016, the Obama Federal Communications Commission passed a broadband privacy rule designed to limit ISPs' ability to monitor customer communications;[30] however, that rule was subsequently rescinded by Congress in 2017.[31] Several states have seen legislation introduced to place similar limitations on ISPs, and in June of 2019, Maine became the first state to pass such restrictions into law.[32]

5. Web Infrastructure

Beyond providing internet service, large internet platforms provide web infrastructure services that may give them insight into what consumers do on other websites. Amazon, for example, provides Amazon Web Services which hosts websites and cloud

[26] Historically, ISPs had not used subscriber data for advertising purposes, but in recent years many of the large ISPs began to build the capacity to monetize personal user data. Matt Keiser, *For Telecoms, The Adtech Opportunity is Massive*, eMarketer (Jan. 18, 2017), https://www.emarketer.com/Article/Telecoms-Ad-Tech-Opportunity-Massive/1015052; *see* Anthony Ha, *Verizon Reportedly Closes in on a Yahoo Acquisition with a $250M Discount*, TechCrunch (Feb. 15 2017), https://beta.techcrunch.com/2017/02/15/verizon-yahoo-250-million/.

[27] Garett Sloane, *Inside Verizon's Plan to Seal Off Its Data (and Conquer Advertising)*, Digiday (Sept. 11, 2015), https://digiday.com/media/inside-verizons-plan-seal-off-data-conquer-advertising/.

[28] Vindu Goel, *Verizon Completes $4.48 Billion Purchase of Yahoo, Ending an Era*, N.Y. Times (June 13, 2017), https://www.nytimes.com/2017/06/13/technology/yahoo-verizon-marissa-mayer.html; Ingrid Lunden, *AT&T Confirms it is Buying Ad Platform AppNexus, Reportedly for between $1.6B-$2B*, TechCrunch (June 25, 2018), https://techcrunch.com/2018/06/25/att-appnexus/.

[29] *FTC Seeks to Examine Privacy Practices of Broadband Providers*, Fed. Trade Comm'n (Mar. 26, 2019), https://www.ftc.gov/news-events/press-releases/2019/03/ftc-seeks-examine-privacy-practices-broadband-providers.

[30] *2019 Privacy Legislation Related to Internet Service Providers*, Nat'l Conference of State Legislatures (June 17, 2019), http://www.ncsl.org/research/telecommunications-and-information-technology/2019-privacy-legislation-related-to-internet-service-providers.aspx.

[31] Brian Fung, *Trump has Signed Repeal of the FCC Privacy Rules. Here's What Happens Next.*, Wa. Post (Apr. 4, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/04/04/trump-has-signed-repeal-of-the-fcc-privacy-rules-heres-what-happens-next/.

[32] Maggie Miller & Emily Birnbaum, *Maine Governor Signs Bill Banning Internet Providers from Selling Consumer Data Without Consent*, The Hill (June 6, 2016, 3:48 PM),https://thehill.com/policy/technology/447345-maine-governor-signs-into-law-bill-to-ban-internet-providers-from-selling.

storage for countless other sites and services; however, its privacy FAQ states that Amazon does not use this data for its own purposes without customer consent.[33] Other large platforms provide similar cloud processing services on behalf of other companies as well.

B. The Mobile Ecosystem

A core difference in the mobile ecosystem is that apps generally have greater access to device capabilities than websites. Further, much more data is collected "passively" as users go about their daily lives, often when apps are running in the background. Mobile operating systems directly make available data stored or generated by the device, such as device storage, geolocation, camera, or microphone. Those secondary data elements are discussed below in *Infrastructure, Other Phone Data*.

1. Mobile Apps

As with websites, major internet platforms offer a wide variety of mobile applications that collect a great deal of sensitive information directly from the user, such as health data (fitness trackers), financial data (payment apps),[34] and communications (social media and email). In some ways, mobile apps are architecturally similar to websites—such as heavy use of JavaScript and heavy reliance upon third-party functionality.. On the other hand, there are important differences. For one, mobile apps do not typically rely upon HTML cookies to keep track of users—instead they rely upon device-generated identifiers. For example, the Android platform creates "Instance IDs" for apps to keep state on users from session to session.[35] It also offers a cross-app "Advertising ID" to allow for third-parties to track users across different apps; however, this ID is resettable, and Android has a setting to instruct apps to disallow tracking for ad targeting. Mobile devices also have persistent hardware or system IDs (such as the "Android ID"), but applications are generally dissuaded from using them for tracking.

2. Ad Tech Tracking Companies

Also with websites, apps embed numerous third-party technologies that collect data about user behavior in those apps. The mechanism by which these third parties are embedded is slightly different than the web—instead of configuring a site to include tags

[33] "We do not access or use your content for any purpose without your consent. We never use customer content or derive information from it for marketing or advertising." Data Privacy: Overview, AWS, https://aws.amazon.com/compliance/data-privacy-faq/.
[34] Tobie Stanger, *Why Apple Pay is the Highest-Rated Peer-to-Peer Payment Service*, Consumer Reports (Aug. 6, 2018), https://www.consumerreports.org/digital-payments/mobile-p2p-payment-services-review/.
[35] Best Practices for Unique Identifiers, Android Developers, https://developer.android.com/training/articles/user-data-ids.

that initiate connections to different servers, third parties offer "software development kits" (SDKs) that app developers can plug into their code for purposes such as advertising, analytics, and web measurement. Once an SDK is installed, however, it has the ability to transmit data about the app to the third-party company. This data can include incredibly detailed and sensitive information; for example, in one Wall Street Journal investigation, a fertility tracking app was found to be transmitting to Facebook (via a Facebook analytics SDK) information about when a user had her period, or if she was intending to get pregnant.[36]

As with the web, many of the largest internet platforms offer third-party functionality for app developers to use for things like advertising, analytics, social media integration, and single-sign on. These SDKs typically transmit data about the user back to the platforms. In the AppCensus review of 426 platform apps, in the majority of cases those apps were sharing data through third-party SDKs to other platform companies.

3. App Stores, Operating Systems, and Hardware

In the vast majority of cases, users get mobile applications from app stores native to the phone's operating system: the App Store for iPhones and the Google Play Store for Android. In addition to knowing the identity of the applications installed, the operating system also knows when and for how long users interact with those apps. The OS may have the technical means to further monitor what users do in those applications, but it is not always clear from the companies' policies the extent to which they engage in this behavior. Other platforms may also offer their own app stores to mobile users, such as Amazon's store for Fire and other mobile devices. Hardware and OS providers may also preinstall default applications that have the ability to collect data about app usage in other ways, such as virtual keyboard software that transmits information detailing keystrokes and commonly used words.[37]

As with desktop computers, smartphone hardware manufacturers who do not write the operating system have the capacity to install software to monitor device usage—especially on phones running Android which is open-source and designed to be easily configurable. Some Amazon devices, for example, run a heavily customized version of Android, which gives Amazon the ability to collect telemetry data about how the device is used.[38]

---

[36] Sam Schechner & Mark Secada, *You Give Apps Sensitive Personal Information. Then They Tell Facebook.*, Wall St. J. (Feb. 22, 2019, 11:07 AM), https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636.
[37] *Learn How Gboard Gets Better*, GBoard Help, https://support.google.com/gboard/answer/9334583.
[38] The Crow Grandfather (@TheCrowGrandfather), Reddit (Dec. 31, 2017, 2:53 PM), https://www.reddit.com/r/kindlefire/comments/7nal4a/kindle_telemetry_and_battery_drain/.

4. Internet Service Providers

Internet service providers have insight into the mobile traffic that flows through their networks, though encryption may limit what they can tell about the substance of the content. If nothing else, ISPs will often be able to understand the source (the domain) of mobile traffic, as well as the general nature of the traffic as ISPs treat different types of traffic (such as web browsing or video) differently. In some cases, researchers have alleged that mobile carriers specifically throttle certain video traffic, such as traffic from YouTube and Netflix.[39] Carriers who sell phones directly through their stores also may insist on the installation of custom software which may facilitate additional data collection.[40]

5. Infrastructure

As with the web, large internet platforms may provide hosting, cloud, or related services for other applications. By providing those services to other companies, the platforms may have insight into what consumers do through those applications.

6. Geolocation

With the advent of mobile phones, a host of companies now get access to detailed information about users' movements over the course of the day. Device geolocation can be generated in a number of different ways[41] and is widely shared among companies. Geolocation data is typically considered to be extremely sensitive by regulators, as it can reveal intimate details about an individual's life, can compromise physical safety if obtained by bad actors, and can be uniquely identifying even when not otherwise tied to real-world identity.

a) Internet Service Providers

When turned on, mobile phones constantly look for, and connect to the closest cell phone towers in order to provide service. As such, mobile service providers have access to a detailed history of all the places we travel to while our smartphones are on. Generally speaking, these providers make few statements about how long that data is retained or limit how that data will be used to make inferences about users—though they typically do offer various opt-out controls around the use of this data for advertising

[39] Kendra Chamberlain, *Study Claims Streaming Video Apps are Being Throttled by Wireless Carriers*, Fierce Wireless (Sept. 5, 2018, 12:33 PM), https://www.fiercewireless.com/wireless/study-claims-streaming-video-apps-are-being-throttled-by-wireless-carriers.
[40] Shaun Nichols, *Eggheads Confirm: Rampant Android Bloatware a Privacy and Security Hellscape*, The Register (May 9, 2019, 7:05 PM), https://www.theregister.co.uk/2019/05/09/android_bloatware_security/.
[41] Statement of Justin Brookman, *Mobile Privacy: Your Smartphones, Tablets, Cell Phones, and Your Privacy Hearing*, Senate Judiciary Comm. Subcommittee on Privacy, Tech. & the Law (May 10, 2011), https://cdt.org/files/pdfs/20110510_mobile_privacy.pdf.

and consumer research.[42] ISPs have also shared data with third-party aggregators, who in turn have sold precise geolocation to a host of other third parties, including bounty hunters and jealous lovers.[43] Of course, both mobile and fixed ISPs also generally know consumers' home addresses for billing purposes, at a minimum.

b) Mobile Operating Systems and Hardware

Mobile operating systems also regularly generate geolocation data for use in mobile applications (such as maps and navigation), though often using different methods than mobile carriers. Originally, cell phones used traditional GPS technology based on measuring the signals from dedicated geolocation satellites in stationary orbit around the earth. Over time, the OS companies began to supplement this method by building out maps based on the location of persistent WiFi access points (determining geolocation by referencing nearby access points is much quicker and battery-efficient). For this reason, as a condition of turning on location services, OS companies will constantly monitor and report back to the company, nearby wireless access points in order to keep their maps up-to-date. As a result, these companies, like mobile carriers, can collect a detailed map of all the places you take your phone over time. Both Apple and Android offer tools to disassociate location history from identity, though those settings are off by default. Hardware manufacturers (even if they are not the primary developer of the operating system) may also have the capacity to configure phones to generate and transmit geolocation information.

c) Applications and Websites

Mobile applications can request access to geolocation through a phone operating system's location permission, even when those applications are not in use.[44] Due to a number of perceived abuses in geolocation access, mobile operating systems have over time implemented a number of changes to constrain geolocation collection by apps. For example, both Android and iOS display indicators when an app is accessing geolocation. Similarly, both operating systems require a stand-alone grant of permission to the app, in response to a dedicated prompt in order to give an app access to geolocation. Apple has introduced controls to allow applications to only access geolocation when they are actively being used by the user.

---

[42] Recently, several of these providers were found to be selling precise geolocation through middlemen to a variety of third parties, with few checks to prevent abuse. Joseph Cox, *Big Telecom Sold Highly Sensitive Customer GPS Data Typically Used for 911 Calls*, Vice (Feb. 6, 3019, 5:11 PM), https://www.vice.com/en_us/article/a3b3dg/big-telecom-sold-customer-gps-data-911-calls)?

[43] Joseph Cox, *Hundreds of Bounty Hunters Had Access to AT&T, T-Mobile, and Sprint Customer Location Data for Years*, Vice (Feb. 6, 2019), https://www.vice.com/en_us/article/43z3dn/hundreds-bounty-hunters-att-tmobile-sprint-customer-location-data-years.

[44] Jennifer Valentino-DeVries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. Times (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

Nevertheless, applications may be able to access a user's geolocation through other means without querying a phone's permission and triggering consent or notice. Apps have access to the user's IP address, and can in many cases generate rough geolocation just from that—with an increasing level of precision.[45] An app may also be able to access the identity of the wireless hotspot through which a user is connecting in order to cross-reference an external database that maps out hotspots. Or the app may access the operating system's cache of previously generated hotspot MAC addresses without triggering a request for present geolocation through the dedicated geolocation API.[46] In our review of over 400 apps offered by the largest internet platforms, at least 24 apps were transmitting the identity of the wireless hotspot, which may have been a way to determine geolocation.

Websites may also have the capacity to access geolocation on devices that generate such data, though the interface is less seamless, and desktop computers may not have GPS or access to wireless connectivity to generate geolocation.

Even if an app is not accessing geolocation using these methods, it may be able to estimate location just from the user's IP address. This location information is not as precise as location generated from a GPS chip or triangulated from nearby cell towers or WiFi access points, but can still provide a general sense of where the user is. Facebook, for example, tracks users' rough location based on IP address even if a user has configured the app to not collect geolocation from the phone.[47]

d) Ad Tech Tracking Companies

Once an app is granted permission to access geolocation, there are few limitations on what it can do with that data—or with whom the data can be shared. Ad tech SDKs are routinely configured to request geolocation from an app, and in many cases are the sole reason an app requests access to geolocation in the first place. However, recently, Apple has started to remove apps from its App Store that do not explicitly receive permission from the consumer to share location data with third parties.[48]

---

[45] David Bilson, *Finding Yourself: The Challenges of Accurate IP Geolocation*, Oracle DYN (Jan. 29, 2018), https://dyn.com/blog/finding-yourself-the-challenges-of-accurate-ip-geolocation/.
[46] Joel Reardon et al., *50 Ways to Leak Your Data: An Exploration of Apps' Circumvention of the Android Permissions System*, PrivacyCon 2019, https://www.ftc.gov/system/files/documents/public_events/1415032/privacycon2019_serge_egelman.pdf.
[47] Kashmir Hill, *Turning Off Facebook Location Tracking Doesn't Stop It from Tracking Your Location*, Gizmodo (Dec. 18, 2018, 12:20 PM), https://gizmodo.com/turning-off-facebook-location-tracking-doesnt-stop-it-f-1831149148.
[48] Thur Ong, *Apple is Reportedly Removing Apps That Share Your Location Data with Third Parties*, The Verge (May 9, 2018), https://www.theverge.com/2018/5/9/17334602/apple-targeting-apps-location-data-sharing-third-parties.

7. Other Phone Data

Modern smartphones offer a dizzying array of functionality that provides consumers with real benefits, but also presents privacy threats in that personal data may be collected and shared in ways that are difficult to measure.

a) Microphone

Cell phone microphones aren't just limited to use during phone calls. Operating systems are often configured to listen constantly for a "wake word" that initiates a function on the device; although this processing is designed to limit the collection of non-triggering data, in some cases operating systems have been found to collect and transmit conversations and behavior that was never intended to be shared.[49] The operating systems make the microphone available to installed apps as well, and many consumers have wondered whether apps are accessing the microphone to record conversations unrelated to its functionality. Facebook, especially, has been the subject of constant speculation of eavesdropping on users for ad targeting, though the company has adamantly denied these claims.[50]

While some app developers deny accessing phone microphones for ad targeting, some third-party ad tech companies are aggressively trying to do just that. Startups like Alphonso offer SDKs that listen for ambient background noise to try to infer information about the user,[51] and in 2016 the Federal Trade Commission sent warning letters to app developers who embedded code from the company Silverpush, which was configured to listen to users in the background.[52] However, there is no evidence the large platforms that are the subject of this report are using these technologies, either in their own third-party SDKs, or as first-party applications.

[49] Alex Hern, *Apple Contractors 'Regularly Hear Confidential Details' on Siri Recordings*, The Guardian (July 26, 2019, 12:34 PM), https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings; Elizabeth Weisse, *Alexa Creepily Recorded a Family's Private Conversations, Sent them to Business Associate*, USA Today (May 24, 2018, 6:10 PM), https://www.usatoday.com/story/tech/talkingtech/2018/05/24/amazon-alexa-creepily-recorded-sent-out-familys-conversations/642852002/; *Google Pauses Listening to EU Voice Recordings, Probe Begins*, SIFY Finance (Aug. 2, 2019, 3:18 PM), https://www.sify.com/finance/google-pauses-listening-to-eu-voice-recordings-probe-begins-news-finance-ticpsJfbajdgh.html.

[50] Kaitlyn Tiffany, *The Perennial Debate About Whether Your Phone is Secretly Listening to You, Explained*, Vox (Dec. 28, 2018, 11:50 AM), https://www.vox.com/the-goods/2018/12/28/18158968/facebook-microphone-tapping-recording-instagram-ads.

[51] Sapna Maheshwari, *That Game on Your Phone May Be Tracking What You're Watching on TV*, N.Y. Times (Dec. 28, 2017), https://www.nytimes.com/2017/12/28/business/media/alphonso-app-tracking.html.

[52] *FTC Issues Warning Letters to App Developers Using 'Silverpush' Code*, Fed. Trade Comm'n (Mar. 17, 2016), https://www.ftc.gov/news-events/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code.

b) Bluetooth

Most phones have the ability to communicate via Bluetooth, a wireless technology standard that allows relatively short-distance interactions between devices. Bluetooth allows easy usage of peripherals such as external speakers, but can also facilitate tracking in physical spaces that have Bluetooth tracking beacons installed. Smartphone apps can be configured to passively scan for nearby beacons and to communicate back to the app developer whenever one is encountered.[53] Even if a user does not install an app with access to Bluetooth, smartphones with WiFi enabled constantly broadcast information in order to easily connect to previously-encountered hotspots. Companies that install physical sensors can generate a longitudinal view into when and how a particular device visits a unique location.[54]

c) Contact Information

Carriers process phone calls and SMS messages, and can typically see with whom individuals are communicating. Smartphone operating systems also generate a log of the individuals that a person calls and allow users to persistently store contact data, which is typically backed up to the OS's cloud (they may also provide a phone's default messaging app). Mobile apps can request permission to access a mobile device's contacts. Other services, such as Facebook, provide stand-alone messaging applications for consumers to download. Even setting aside the content of all these communications, contact information can be extremely revealing, and companies may try to derive socioeconomic or interest data from this information, or to expand their user base by suggesting acquaintances to add to a social networking service.[55]

d) Environmental Sensors

Smartphones increasingly include sensors such as accelerometers and barometers, which can be used to make various inferences about how and where a user is accessing the device. Google, for example, uses this data to assess whether a user is likely walking, biking, or riding a train.[56] These sensors can also be used to deduce

[53] Christopher McFadden, *Are You Being Tracking by Bluetooth Beacons while Shopping?*, Interesting Eng'g (June 20, 2019), https://interestingengineering.com/are-you-being-tracked-by-bluetooth-beacons-while-shopping.
[54] Neal Ungerleider, *Google Analytics for Real Life: Tracking Retail Customers Through Smartphones*, Fast Co. (Jan. 14, 2013), https://www.fastcompany.com/3004781/google-analytics-real-life-tracking-retail-customers-through-smartphones.
[55] Kashmir Hill, *'People You May Know:' A Controversial Facebook Feature's 10-Year History*, Gizmodo (Aug. 8, 2018, 8:25 AM), https://gizmodo.com/people-you-may-know-a-controversial-facebook-features-1827981959.
[56] Google Data Collection, Digital Content Next (Aug. 2018), https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf.

health data about an individual, such as how many steps a person takes (at least while carrying their smartphone) during the course of a day.[57]

Other device-level permissions that smartphones provide include:

- Device storage;
- Camera and photos;
- Calendar; and
- Call and SMS logs.

All of this data may be available to apps and other third-parties, not to mention the underlying operating system vendor (who often offers cloud storage of this data for device backup).

C. Television Viewing

Even in the age of ubiquitous mobile devices, televisions remain the screen that consumers look at the most.[58] Historically, data about what we watch on our televisions has been viewed as sensitive and worthy of federal statutory protection. The Cable Act imposes limitations on what cable providers can share about viewing, and the Video Privacy Protection Act prohibits the sharing of data about video rental records. However, as televisions become "smarter," manufacturers and other companies are increasingly able to monitor and share data about viewing habits to more closely approximate the sophisticated data sharing environment for web browsing and mobile app usage.[59]

1. Cable Service Providers

First, your cable or internet service provider has the capacity to monitor viewing behavior by virtue of delivering the service. Traditional cable services are able to precisely monitor the shows watched over time, which can be used to generate a detailed profile about a household's preferences. Service providers may have more difficulty monitoring viewers who watch traditional television content (or other video) through the web or standalone applications; however, as with other web content (*see supra*, The Web Ecosystem, Internet Service Providers), they will at least be able to

[57] Chris Hoffman, *How to Track Your Steps with Just an iPhone or Android Phone*, How-To Geek (July 20, 2016, 1:44 AM), https://www.howtogeek.com/238904/how-to-track-your-steps-with-just-an-iphone-or-android-phone/.
[58] *People Spend Most of Their Working Hours*, *supra* note 9.
[59] Smart TV Workshop Slides, Fed. Trade Comm'n (Dec. 7, 2016), https://www.ftc.gov/system/files/documents/public_events/942763/smarttv_workshop_slides.pdf.

discern the sites or apps used, which could be indicative of content or otherwise illuminating.

2. Televisions, Operating Systems, and Software

Newer smart televisions increasingly have the ability to monitor wihch programs are being watched and report back to the manufacturer or various embedded third parties. A Consumer Reports investigation in 2018 demonstrated that several major manufacturers were using a variety of technologies to discern what content was being displayed on the televisions' screens.[60] Most notably, all the televisions used "automated content recognition" (ACR) to regularly send screenshots of content back to a service provider to compare against a comprehensive library of images to match up against known programming.[61] In 2017, the Federal Trade Commission brought an action against Smart TV manufacturer Vizio alleging that Vizio collected and shared viewing behavior without consumers' affirmative permissions.[62] However, the interfaces that Smart TVs use in order to collect such permission are not always clear and understandable.[63]

Peripheral devices that enable streaming viewing on other devices, such as the Amazon Fire Stick, have the capacity to monitor and collect viewing behavior.[64] Other hardware, such as computers and laptops, also has the technical ability to capture images and send data to remote servers for processing and investigation, but such behaviors have not as yet been documented.

In some cases, a device manufacturer itself may not track viewers, but it embeds third-party software—either the television's operating system or otherwise—that does. Some manufacturers, for example, embed software from Samba TV that allow them to generate rich profiles of viewing behavior for ad targeting, measurement, or other

---

[60] *Samsung and Roku Smart TVs Vulnerable to Hacking, Consumer Reports Finds*, Consumer Reports (Feb. 7, 2018), https://www.consumerreports.org/televisions/samsung-roku-smart-tvs-vulnerable-to-hacking-consumer-reports-finds/.

[61] Smart TV companies can also track what channels are viewed on the TV, and can cross-reference local cable listings to determine what shows are watched; however this data gathering is limited to the content viewed on traditional television channels; ACR gives companies the ability to monitor viewing on Netflix and other over-the-top services, though more fringe, long-tail videos (such as those generated by users) may not be included in the company's catalog of recognizable content.

[62] *Vizio to Pay $2.2 Million to FTC< State of New Jersey to Settle Charges it Collected Viewing Histories on 11 Million Smart Televisions without Users' Consent*, Fed. Trade Comm'n (Feb. 6, 2017), https://www.ftc.gov/news-events/press-releases/2017/02/vizio-pay-22-million-ftc-state-new-jersey-settle-charges-it.

[63] James K. Wilcox, *How to Turn Off Smart TV Snooping Features*, Consumer Reports (Oct. 24, 2018), https://www.consumerreports.org/privacy/how-to-turn-off-smart-tv-snooping-features/.

[64] *Stop Amazon from Spying on You*, TroyPoint, https://troypoint.com/stop-amazon-from-spying-on-you/.

purposes.[65] Several televisions run on the Android operating system managed by Google; as part of setting up the television, viewers are required to agree to Google's Terms of Service and Privacy Policy.[66] However, due to the breadth and vagueness of Google's privacy policy, it is difficult to tell how actively Google monitors viewing behavior on televisions apart from keeping track of the apps installed through the app store.

3. Apps

Just as websites and mobile apps can collect first-party data on what users do on their sites, Smart TV apps like Amazon and YouTube can monitor what content consumers view through those services. As with apps, Smart TV operating systems may provide access to unique identifiers which allow consumers to be tracked more effectively over time. In many cases, consumers must log into those services in order to access the content, but even when they don't, the service may be able to correlate Smart TV app behavior to behavior on other devices by matching an IP address or other shared traits (see *infra*, How Companies Tie All This Data Together).

4. Ad Tech Ecosystem

Some television operating systems are moving to standardize this data collection and sharing by making persistent device identifiers available to apps and their partners—including dedicated advertising identifiers.[67] These identifiers allow tracking companies to keep state on users over time and to track behavior across different smart TV applications. Although these advertising identifiers are often resettable and/or accompanied by settings to limit tracking, the default in many if not most cases is typically to allow tracking; it is unclear how many consumers are aware of or take advantage of smart TV privacy controls.

In addition to smart TV apps sharing data with third parties, other entities, such as the manufacturer, may also be sharing data with third parties advertising companies. As part of its investigation into smart TV maker Vizio, the Federal Trade Commission alleged that the company was sharing information about television viewing behavior

[65] Sapna Maheshwari, *How Smart TVs in Millions of U.S. Homes Track More than What's On Tonight*, N.Y. Times (July 5, 2018), https://www.nytimes.com/2018/07/05/business/media/tv-viewer-tracking.html.
[66] Katie McInnis, *Evaluation of the Privacy and Security Aspects of Connected Televisions*, PrivacyCon 2018, https://www.ftc.gov/system/files/documents/public_events/1223263/panel019_smarttvs_privacy.pdf.
[67] Phil Nickinson, *Roku Privacy Policy and Ad-Tracking: What You Need to Know*, Cordcutters (Aug. 15, 2018), https://www.cordcutters.com/roku-privacy-policy-and-ad-tracking-what-you-need-know.

with third parties tracking companies for audience measurement, ad effectiveness, and to target ads on *other* devices based on a consumers' television viewing habits.[68]

D. The Internet of Things

While Smart TVs have become means of collecting and sharing information about consumer behaviors, other Internet of Things (IoT) devices also allow for extensive first-party data collection, and potentially sharing with third-parties as well. IoT devices are particularly interesting for data-collection because they often involve several components out of necessity to operate. Many IoT devices strive to have minimal surface area for user interaction, letting their connectedness blend into functionality consumers would readily expect from a refrigerator, a light switch, a doorbell, or otherwise. Instead, typically, the execution of an IoT device actually relies on a cluster of related products and services, such as cloud infrastructure, smartphone apps, and third-party app constellations. The potential for data collection and cross-device identification is thereby magnified, allowing a platform to gather information about your identity, your personal preferences, your physical environment, and your devices, by administering the service of an IoT device.

1. Digital Assistants

Digital assistants allow consumers to seamlessly interact with a smart service by issuing voice commands instead of manually entering them on a keyboard. Much of the data collected is similar to other first-party applications, including search terms, shopping, and content preferences. Like smartphone operating systems, digital assistants often have microphones (and sometimes cameras) that are triggered by a specific "wakeword," though in some instances, sensors have processed and transmitted data without the wakeword having been uttered.[69] In a collaborative effort with researchers from Northeastern University in 2019, CR was able to determine a dictionary of quasi-wakewords that caused a variety of digital assistants to wake.[70]

2. Routers

Since routers are the key to the internet for the majority of home and apartment dwellers, they are the conduit through which most (if not all) sensitive data flows. This imposes a heavy amount of trust on routers and router manufacturers, but, thankfully, the recent advent of certificate-based HTTPS traffic (and other encryption schemes) has

[68] Complaint for Permanent Injunction and Other Equitable and Monetary Relief, Fed. Trade Comm'n v. Vizio, Inc.,(Feb. 6, 2017), *available at* https://www.ftc.gov/system/files/documents/cases/170206_vizio_2017.02.06_complaint.pdf.
[69] *See supra* note 48.
[70] Allen St. John, Smart Speakers That Listen When They Shouldn't, Consumer Reports (Aug. 29, 2019), https://www.consumerreports.org/smart-speakers/smart-speakers-that-listen-when-they-shouldnt/.

made wide-scale data collection difficult for routers (and, in fact, any intermediary device) in general. However, routers often have access to traffic and device metadata that can reveal information about the activity of the people using them, and the connected devices they (and often their friends and relatives) own.

Many router manufacturers also integrate malware detection tools, which necessitates the sharing of potentially sensitive data with service providers, such as Trend Micro or McAfee. And, while proper disclosure around this practice is at times weak or non-existent, some manufacturers describe what information they collect, and what they do with it (and, more importantly, what they do not do with it) in more detail.

Initial examinations of the most popular consumer routers by Consumer Reports in 2019 found no explicit transmission of sensitive personal data by routers themselves, however, disclosures from companies like Eero and Google suggest that information is collected about networks and their devices for specific purposes, like benchmarking. (Google states, "We do not share your personal information from your Wifi point or the Google Wifi app for the purposes of advertising without your consent.") Also, it remains to be determined whether each router's associated smartphone app makes use of the common components of the advertising and tracking ecosystem that may otherwise leak personal data.

Many consumer routers are provided directly by Internet Service Providers (ISP). They are similar—if not identical—in construction to other store-bought routers, but may contain firmware altered or built by the ISP who provided the router. While it is possible that these routers could harvest more metadata to feed back to their respective ISPs than the ISPs could gather on their own, Consumer Reports need a consistent way to determine whether or not this activity occurs and to what degree.

3. Home Security Cameras

Home security cameras can collect very sensitive data about individuals, storing that data in the cloud for as long as the user has an account, or perhaps indefinitely. Some companies use artificial intelligence techniques to identify the objects captured by camera footage, though as with digital assistants, companies also use human contractors to manually review and tag surveillance footage.[71] Companies also increasingly process footage from these and other sources for *facial recognition*, which can allow a system to detect an unknown visitor, but can also allow inferences to be made based on a specific individual's patterns of behavior. Some home security cameras' policies include language limiting company access to secondary usage of

[71] Jack Morse, *Ring's 'AI' Reportedly Involved People in Ukraine Watching Customer Videos*, Mashable (Jan. 10, 2019), https://mashable.com/article/amazon-ring-ai-privacy-videos/.

customer data, which Consumer Reports considered in our ratings of 14 home security cameras for privacy and security attributes in 2019,[72] however Amazon, Google, and Facebook's patent filings have at least explored the possibility of using this data in a commercial context. Other IoT devices may be embedded with cameras as well, which could be useful for functionality, but which could also be used to derive data for sale to third parties.[73]

4. Cars

Cars increasingly have the capacity to collect and report back to the manufacturer detailed information about users, including geolocation data, driving behaviors, and data generated from sensors such as cameras and microphones. Google's parent company Alphabet has a self-driving cars division (Waymo) though its leadership has taken pains to distance itself from the data collection practices of Google.[74] Cars also have the capability to share data with and through other third-party platforms; many cars encourage Apple and Android integration with device infotainment systems, and they often rely upon cellular service for connectivity.

5. Gaming

Some large platforms offer dedicated gaming hardware, such as Microsoft's Xbox and Facebook's virtual reality system, Oculus. In addition to data about gameplay, these platforms often facilitate social interactions, or may allow other media viewing or internet usage. However, due to the demand for this technology to offer a realistic virtualized world (one that you can bump into, for example), these platforms also have the ability to map out physical space and identify objects within it.

6. Other IoT Devices

Even single-purpose objects such as smart thermostats and sleep monitors may generate sensitive data, such as when consumers are at home or not. Even if this data is encrypted, anyone with the capacity to monitor those devices' transmissions through

[72] Daniel Wroclawski, *Best Wireless Home Security Cameras of 2019*, Consumer Reports (Jan. 1, 2019), https://www.consumerreports.org/wireless-security-cameras/best-wireless-home-security-cameras-of-the-year/.
[73] Maggie Astor, Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared, N.Y. Times (Jul. 25, 2017), https://www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html.
[74] Rachel Withers, *Waymo Wants You to Know it has No Interest in Your Data*, Slate (Apr. 3, 2018), https://slate.com/technology/2018/04/waymo-wants-you-to-know-it-has-no-interest-in-your-data.html.

the local network may be able to learn about consumer behavior just by analyzing the patterns of data transmissions.[75]

E. Offline Shopping and Other Activity

In addition to the vast amounts of online data that tech companies collect, they also have the ability to supplement that data with data secured from other companies.[76] Companies need only be able to match an identifier, such as an email address or phone number, in order to share information about what consumers do in different contexts.[77] Facebook's Custom Audiences product, for instance, allows other companies to upload lists of individuals they wish to target with Facebook ads.[78] Cable television providers increasingly have the ability to customize ads to individual households; advertisers are able to upload lists of customers they want to reach and the cable provider can serve a targeted ad for matched customers. Last year, it was reported that Google was receiving credit card transactions from Mastercard to help determine when Google ads successfully resulted in a purchase.[79] In many cases, companies allege that these information exchanges are "hashed"[80] or "anonymized," though they do not always offer

[75] Noah Apthrope, Dillon Reisman, & Nick Feamster, *A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic*, Workshop on Data & Algorithmic Transparency (May 18, 2017), https://www.ftc.gov/system/files/documents/public_comments/2016/10/00022-131586.pdf.

[76] Of course, several of these companies also offer their own physical stores from which they can gather data about purchases or derived from cameras and other sensors.

[77] Even if two companies have different identifiers about an individual — say, two different email addresses — about an individual, they still may be able to match those data sets through a third-party data broker who knows that both identifiers are linked to the same individual. Kalev Leetaru, *What Does it Mean for Social Media Platforms to "Sell" Our Data?*, Forbes (Dec. 15, 2018), https://www.forbes.com/sites/kalevleetaru/2018/12/15/what-does-it-mean-for-social-media-platforms-to-sell-our-data/#15615efb2d6c.

[78] About Custom Audiences from Lists, Facebook Bus., https://www.facebook.com/business/help/341425252616329. Other ad platforms offer similar capabilities: *see*, *e.g.*, Create a Customer List, Google Ads Help, https://support.google.com/google-ads/answer/6276125?co=ADWORDS.IsAWNCustomer%3Dfalse&hl=en; Tailored Audiences from Lists, Twitter Bus., https://business.twitter.com/en/help/campaign-setup/campaign-targeting/tailored-audiences/TA-from-lists.html.

[79] Mark Bergen & Jennifer Surane, *Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales*, Bloomberg (Aug. 30, 2018, 3:43 PM), https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales.

[80] A hash is a one-way mathematical function that turns any amount of data into a different fixed length value. The hash value has no clear connection to the input, so it is difficult, given only the hashed output, to reverse the hash to find the original value. However, every time a hash of any single input is calculated, the output is the same. Hashing algorithms (such as MD5 and SHA-256) are designed to avoid "collision" — that is, two different inputs resulting in the same output. Thus, if two hashed outputs match, it is highly likely that the inputs were the same as well. As an example, the MD5 hash of the identifier "user@domain.com" will always be "cd2bfcffe5fee4a1149d101994d0987f." For more discussion on hashing — as well as the practical limitation of hashes — *see Ed Felten, Does Hashing Make Data "Anonymous"?, Fed. Trade Comm'n (Apr. 22, 2012, 7:05 AM),* https://www.ftc.gov/news-events/blogs/techftc/2012/04/does-hashing-make-data-anonymous.

clear details about what technical or policy measures are in place to limit information leakage or misuse.

These companies can also purchase data from one of hundreds of third-party data brokers that sell detailed records about consumer behaviors.[81] Companies often use these services to append demographic data to customer records, but potentially can purchase more detailed information such as interest categories, shopping behavior, and data culled from public records (such as arrest and mortgage records). Our review of the platforms' privacy disclosures revealed little concrete information about whether and how these companies supplemented their databases of individuals with data from external data brokers.

Finally, a lot of offline behavior may be indirectly observed in other ways. Companies that have access to geolocation data may know what stores or other establishments an individual visits—advertising companies are eager to track this if for no other reason than to demonstrate that their messaging is effective. If a company sends you an email confirming a purchase, your webmail company now has information about your purchasing behavior that it may use for other purposes.[82]

F. How Companies Tie All this Data Together

As discussed in the previous section, tying data sets together is relatively straightforward when different sets are linked to the same real-world identifier. Many of the elements discussed above can be tied to a persistent real-world ID such as a name or email address. ISPs have contact information in order to bill you. Operating systems often ask you to create or share an email address, and may collect billing data for app store (and other) purchases. Websites may encourage (or require) you to log in in order to function.

Not all of the data collected by platforms about users is easily linkable to real identity, but large internet platforms have, over time, architected their systems to make this linkage easier. Google has merged its various properties (such as search and email) to tie data to identity information provided for Gmail or Android. Historically, third-party behavioral ad tracking was not tied to real-world identifiers, but Google and Facebook

[81] *Data Brokers: A Call for Transparency and Accountability*, Fed. Trade Comm'n (May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf; John Eggerton, *Senate Data Brokers Report: Huge Data Collection Under 'Veil' of Secrecy*, Multichannel News, (Dec. 18,2013), https://www.multichannel.com/news/senate-data-broker-report-huge-data-collection-under-veil-secrecy-288772.

[82] *Google Still Keeps a List*, *supra* note 11.

now associate this data with login identity, at least when you are actively logged into their first-party services.[83]

Even without access to a unique cross-service identifier, companies may be able to make informed guesses to match users to otherwise unauthenticated data. If two devices share the same IP address or location for significant periods of time, there is a greater likelihood that they are operated by the same individual. If those devices share similar behavioral patterns (such as browsing the same websites), a company may be able to assert more confidence in the likelihood of linkage.[84]

The volumes of data collected by the platforms we examined provide the opportunity to infer information about consumers that is not provided directly. In some cases, the predictive power behind these inferences requires aggregate data on a scale that only the platforms we have studied can muster. Others only require data from a single person, but deep knowledge of individual data points. For instance, the steady stream of temporal and spatial data provided by the geolocation features on a smartphone can be used to establish probable socioeconomic status, religion, political affiliation, sexual preferences, health conditions, and more. Search logs can indicate mental health problems. Previous purchases can indicate fitness level and predict pregnancy.

Unfortunately, due to the technological and economic advantage of the predictive power that inference provides, companies are not forthright with the way these systems work, when they are invoked, and what kind of data they generate. The glimpses we have into this world come in the form of patent filings, and brochures from advertising agencies that want to flaunt their advanced audience segmentation techniques in order to attract the best customers.

IV. Conclusion

The past few years have seen a marked turn in the public's understanding of the threats posed by the widespread collection and use of personal data. But enabling transparency and choice in the digital marketplace is terrifically difficult without public interest research that can keep pace with changes in companies' practices. There is a fundamental asymmetry in knowledge and resources that puts civil society at a disadvantage when the evaluating functionality, claims, and effects of the technology industry — and innovative approaches will be required to structure and maintain an effective consumer protection regime in the 21st century.

[83] *Google Dropped Ban*, *supra* note 18.
[84] *Cross-Device Tracking*, *supra* note 16.

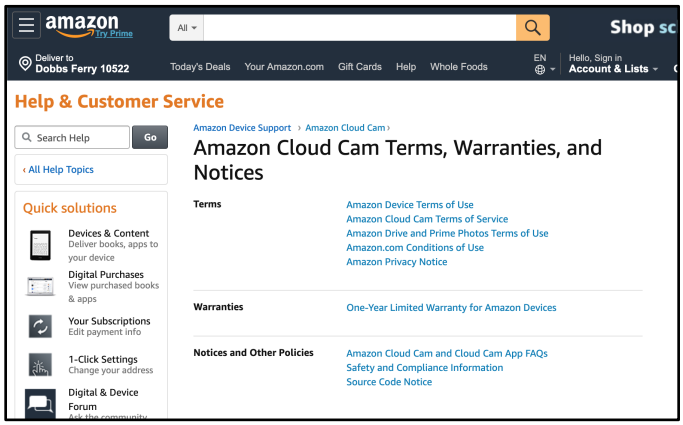
Appendix I: A Case Study of the Amazon Cloud Cam

The Amazon Cloud Cam puts the complexity of the digital marketplace into stark relief. A Cloud Cam is not merely a hardware appliance: it is a service made possible by an accompanying mobile app, a cloud storage system, and Amazon's extensive web infrastructure.

When a user buys an Amazon Cloud Cam, how do they know where their data is going? How do they know what's collected from the device? The app? How do they know where that data is stored? How do they know that it's secure? Or how it's used? Or how it's shared? We found 10 relevant documents that govern a consumer's use of the Amazon Cloud Cam and its accompanying mobile application:

1. Amazon.com Privacy Notice
2. Amazon.com Conditions of Use
3. Amazon Device Terms of Use
4. Interest Based Ads @Amazon
5. EU-US Privacy Shield
6. Amazon Cloud Cam Terms of Service
7. Report a Security Issue
8. Amazon Cloud Cam & App FAQs
9. Amazon Photos Terms of Use
10. File Retention Policy

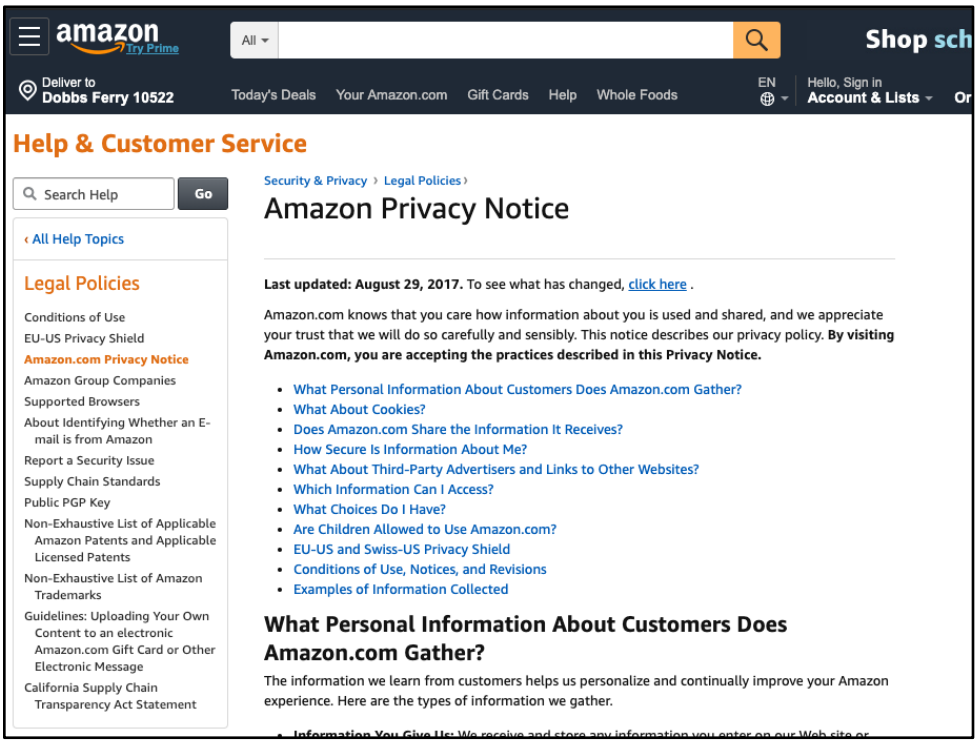
How would a consumer begin to understand the terms and conditions underlying their use a CloudCam? How might this consumer be confident that she had consented to Amazon's data practices?



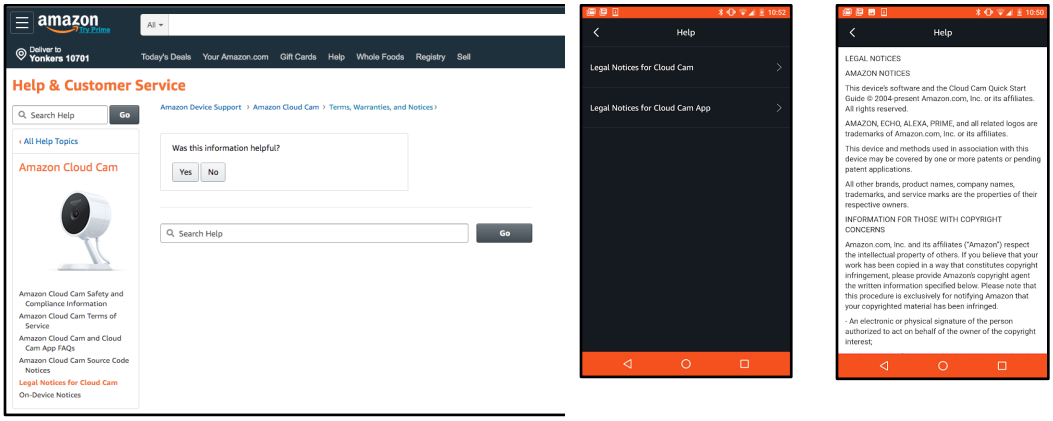
Users can access these documents from multiple locations: the Amazon Cloud Cam microsite, Amazon.com, hyperlinks within other documents, and the Cloud Cam app.

Understanding which documents apply is difficult: 6 out of 10 documents can be found under the Support page for the device under the “Terms, Warranties and Notices” link; the “Amazon Drive and Prime Photos Terms of Use” link has a separate title for the document labeled “Amazon Photos Terms of Use”; the File Retention Policy can be found through a hyperlink from the Amazon Photos Terms of Use.

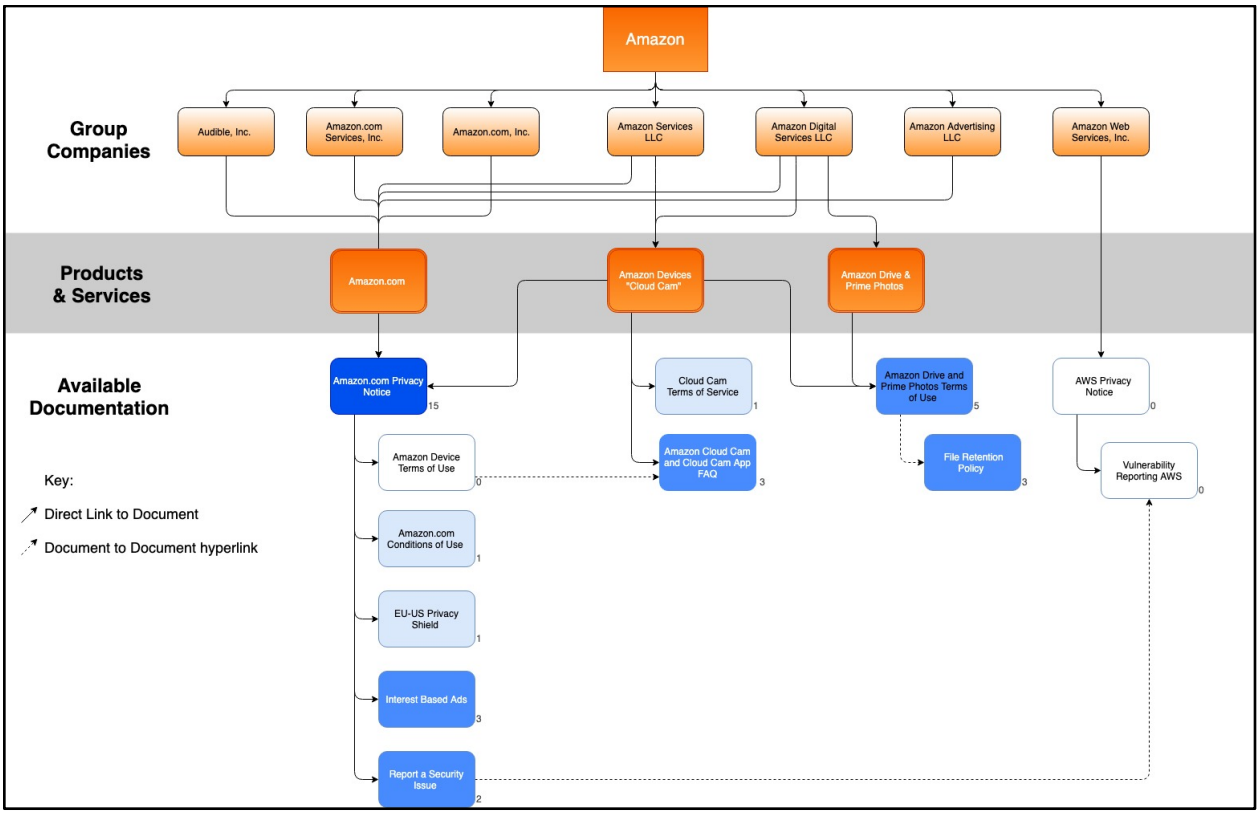
When directed to the Amazon.com Privacy Notice, all of the additional documents can be found under the Legal Policies tab on the left hand side, including those which cover Amazon’s subsidiaries. The “Interest-Based Ads” Policy can be found as a hyperlink in the Amazon.com Privacy Notice, or at the bottom of every amazon.com webpage.

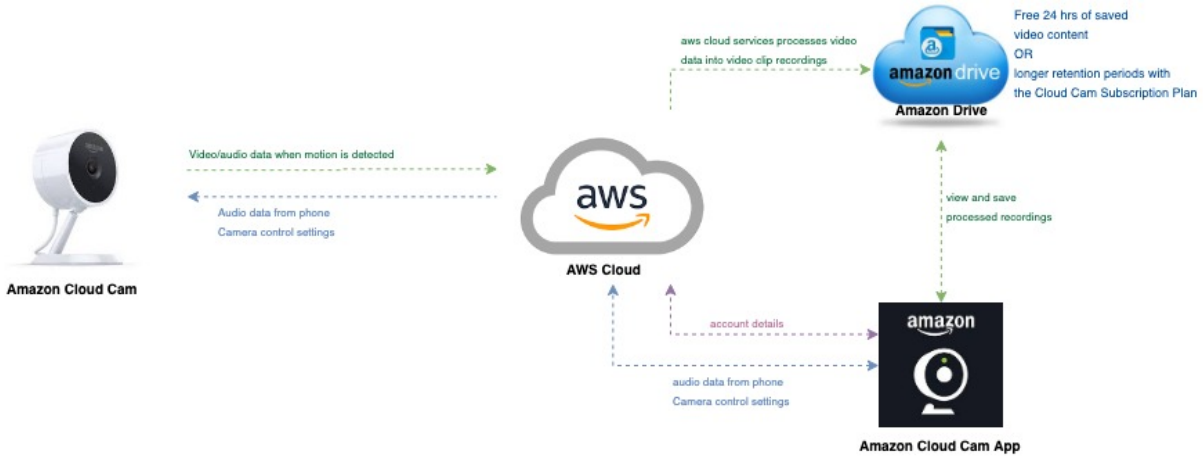


Moreover, certain legal notices can be found within the Cloud Cam app, but not on the website:



The Amazon Device Terms of Use classifies the Cloud Cam as a "Amazon Device," for which the Terms of Use falls under Amazon Digital Services LLC. We might use Amazon Cloud Cam services, or Amazon Drive (or Photos), the use of which would be governed by a separate agreement, under a separate group company:





From a consumer standpoint, there are too many documents that relate to the product and not enough coverage in them. The privacy notice is vague on critical points such as data storage and usage. The app gives some information on data storage, but if it can only be found on the app, then it cannot be put into a document review or automatically monitored for changes.

All of this makes it very challenging for consumer protection organizations like CR to maintain meaningful, ongoing privacy and data security ratings within frameworks like the Digital Standard. More to the point, it illustrates the challenges facing consumers when choosing products and services that align with their expectations.

CR has always confronted disparities in market power, consumer choice, and individual control from the same crucial position: before you can fix something, you need to understand it. For this reason we are investigating how best to structure and maintain a knowledge base about companies' data practices, which can support a broader community of researchers, advocates and storytellers.