Written Testimony of Markham Erickson
Vice President of Government Affairs and Public Policy, Google

Senate Committee on the Judiciary
Subcommittee on Competition Policy, Antitrust, and Consumer Rights

"Big Data, Big Questions: Implications for Competition and Consumers"
September 21, 2021

Chairwoman Klobuchar, Ranking Member Lee, and distinguished members of the Subcommittee:

Thank you for the opportunity to appear before you today. My name is Markham Erickson. I am a Vice President of Government Affairs and Public Policy at Google. I have worked at the intersection of technology and policy for over 25 years, primarily in private practice where I had the opportunity to work on issues that defined the evolution of the modern Internet. In my role at Google, I oversee a global team of subject matter experts focused on the application of law and policy to technology and the internet, including on issues related to data governance and competition.

We appreciate the topic of this hearing on the implications of data for competition and consumers. Data should be used to make consumers' lives better by improving the quality and diversity of products and services available,[1] while protecting users' privacy and giving them control over their data. In this testimony, I will describe (1) how Google uses and protects data; (2) how data mobility empowers consumers and boosts competition, including data portability and advancing open data; and (3) that data alone does not guarantee better products for consumers.

How we use and protect data

Data plays an important role in making the Google products and services people use everyday functional and helpful. We are committed to treating that data responsibly and protecting privacy with strict protocols and innovative privacy technologies.

[1] For example: Doctors can use data to improve patient outcomes, scientists rely on data to better predict weather, and governments rely on data to monitor things like the impact of the COVID-19 pandemic. And brick and mortar retail businesses reduce costs and connect with and help consumers through the innovative use of data.

Google combines insights from data with industry-leading technology to develop products that help people find directions, grow their businesses, or search for information. Google provides information about the personal data we collect and how we use it—such as to provide better services—to our users and the public. We provide notice not just through our privacy policy[2], but also through our work to actively inform individuals about data use in the context of the services themselves, helping to make the information relevant and actionable. On an individual level, what data is collected and how it is used depends on how each person uses our services and how they manage their privacy controls.

Data is one element of our advertising business, where it helps us connect people with relevant advertisements. Advertising is Google's main source of revenue, and enables us to make many of our flagship products, including Search and Maps, available for free to billions of people around the world.[3] Google's advertising platforms support businesses of all types and sizes by helping them reach customers. To give an example of what this means for small businesses, the owner of a local flower store, Studley Flowers in Rochester, New Hampshire said, "Google Ads has provided us with a cost-effective way to compete with the national flower delivery brands and their larger ad budgets. Additionally, it provides us with a way to make sure that when locals search for flowers, we stay top of mind against those larger brands. Our investment in Google tools has certainly paid off, and we're excited to see what's next."[4]

Google also helps millions of website publishers earn advertising revenue on their sites and apps. The ads shown are informed by a search query or page content, but can also be based on a user's interests or other personal data if their privacy settings permit.[5] We do not sell our users personal information to advertisers, or to anyone else.

Our business relies on ensuring our users' trust, specifically in how we use and protect their data. We do this in part by offering industry-leading controls to manage privacy and empowering users to adjust what data is stored in their Google Account. Three billion users visit their Google accounts every year, where they can review and change their privacy settings and delete data stored in their account. For example, Google

---

[2] https://policies.google.com/privacy?hl=en-US
[3] https://about.google/intl/en_US/how-our-business-works/
[4]https://partnertestimonials.withgoogle.com/?_ga=2.19953275.2136290102.1632085855-932915823.1630594247#q[15]
[5] https://howwemakemoney.withgoogle.com/

users can always turn off personalized ads[6] in their Account Settings[7] while still using Google products for free. And since 2019,[8] we have offered industry-leading auto-delete features that give our users the option to automatically delete data from their account, like their Search History. Starting in 2020,[9] new accounts auto-delete some account data by default after 18 months. We regularly prompt our users to review and manage their privacy and security settings with emails and promotions on the Google homepage.

We also invest in research and development of cutting-edge privacy and security engineering techniques that are applied in our products. We share a range of these innovative tools in open source formats, which are free for anyone—including competing companies—to use, benefitting the broader ecosystem. For example, Google's differential privacy library has been used by hundreds of developers around the globe to gain insight from data while protecting individual privacy.[10]

In addition to putting users in control, we keep their data secure by default. Every day, we block 100 million phishing attempts and 15 billion spam messages in Gmail and encrypt four billion photos[11]. Our free Safe Browsing[12] tool helps keep the rest of the Internet secure, automatically protecting more than four billion devices every day by showing warnings to users when they attempt to navigate to dangerous sites or download dangerous files.

We constantly innovate to improve privacy across our own products and on our platforms. This sometimes means finding ways to reduce the amount of personal information needed to achieve similar outcomes. For example, we recently announced Privacy Sandbox, a collaborative initiative that aims to help build a more private and secure web. Through the Sandbox initiative, we are working with the ads industry to develop new digital advertising tools that protect people's privacy and prevent covert tracking, while continuing to support an open and free Internet enabled by advertising. From the start of this project, we developed these tools in an open manner, and sought feedback from industry partners, civil society, and governments. Because many publishers and advertisers rely on online advertising to fund their websites to connect

[6] https://support.google.com/accounts/answer/465?hl=en&co=GENIE.Platform%3DDesktop
[7] https://support.google.com/ads/answer/2662856?
[8] https://blog.google/technology/safety-security/automatically-delete-data/
[9] https://blog.google/technology/safety-security/keeping-private-information-private/
[10] https://developers.googleblog.com/2019/09/enabling-developers-and-organizations.html
[11] https://blog.google/technology/safety-security/our-work-keep-you-safe/
[12] https://safebrowsing.google.com/

with customers, getting this balance right is key to keeping the web open and accessible to everyone. We are proud of our work to apply innovative solutions that balance these interests.

Data mobility empowers consumers and boosts competition

There are considerable options to use available data safely, including tools and resources from Google. Data can be transferred safely in two ways: personal information can be moved from one service to another based on a specific individual's instruction, and datasets can be shared safely as open data using appropriate privacy technologies.

Data portability

Google has been a leader on data portability for over a decade, enabling our users to export their data and take it to another platform. Google's approach to data portability is simple: the user comes first.[13] Data portability empowers consumers to choose services or online platforms based on quality and individual preference—not because they are locked in or because they can not move their data to alternatives.

Since 2011, Google Takeout[14] has allowed users to export their data from over 70 Google products and download them in machine-readable formats so that they can be easily uploaded to another service. Takeout makes it possible for users to move their content to competing services,[15] so no one feels they have to continue using Google if they prefer a service of another company. Since our launch of Takeout, users have exported more than one billion gigabytes from Google products.[16]

Additionally, through our leadership in the Data Transfer Project,[17] Google makes it easier for companies of all sizes to provide tools that let users move data between online services. The Data Transfer Project, a partnership among Google, Apple, Facebook, Twitter, Microsoft, and Smugmug, supports the direct transfer of data between providers, allowing consumers to seamlessly and securely transfer their data directly from one provider to another, rather than downloading and re-uploading their

[13] https://publicpolicy.googleblog.com/2009/09/introducing-dataliberationorg-liberate.html
[14] https://takeout.google.com/u/0/settings/takeout
[15] https://support.google.com/accounts/answer/3024190?hl=en
[16] https://www.regulations.gov/document?D=FTC-2020-0062-0011
[17] https://datatransferproject.dev/

content. The code available through the Data Transfer Project reduces the engineering work any individual company has to do to offer portability to their users.[18]

Data portability benefits both consumers and competition. Giving users control over their data through easy-to-use data export tools boosts competition by reducing the burden of switching services. It paves the way for innovative and new opportunities for service providers of all sizes, and empowers people to try new services and choose the offering that best suits their needs.

Advancing open data and AI for All

Artificial intelligence (AI) is a transformative technology, and we are committed to making it accessible to more people and institutions. Google is a leader in releasing public data and tools that support advanced technological applications like machine learning (ML) and AI.

Many of the largest successes in machine learning have come from data that is openly available on the web. Google has released over 80 labeled datasets for ML researchers and developers to use.[19] For example, we open sourced the Open Images dataset to provide developers with geographically diverse images to support more inclusive results for underrepresented cultures in ML models. And through TensorFlow—an ecosystem of tools, programming libraries, and community resources—we are enabling developers from startups, large companies, and nonprofits to more easily use ML to build a variety of applications, from fraud detection in digital payments to online English grammar correction.[20] TensorFlow has more than 200,000 users and over 160 million downloads, and supports over $15 billion in economic impact in industries including computer manufacturing and software publishing.

In some cases, data that is already publicly available is hard to use because it is hard to find. Google has developed cutting edge tools to address this issue.[21] For example, Google pays for storage of, and provides public access to, public data available through Google Cloud Public Dataset Program.[22] Kaggle makes data science easier by providing all the code needed to leverage over 80,000 public datasets.[23] We also launched a search engine for datasets called Dataset Search.[24] Using a simple keyword

[18] https://www.regulations.gov/document?D=FTC-2020-0062-0010
[19] https://ai.google/tools/datasets/
[20] https://www.tensorflow.org/
[21] https://www.blog.google/technology/ai/sharing-open-data/
[22] https://cloud.google.com/solutions/datasets
[23] https://www.kaggle.com/datasets
[24] https://datasetsearch.research.google.com/help

search, users can discover datasets hosted in thousands of repositories across the Web.[25]

Google also makes insights from our data publicly available in privacy-safe formats, contributing to important research around the world. For example, Google Trends[26] was launched in 2006 and provides access to a largely unfiltered sample of actual search requests made to Google. It is anonymized (no one is personally identified), categorized (determining the topic for a search query), and aggregated (grouped together). Google Trends is a powerful tool for researchers, journalists, and civil society. There are more than 21,000 research papers on Google Scholar that cite Trends as a data source. [27]

During COVID-19, we have leveraged our expertise in open data and privacy to support governments, health officials, researchers, nonprofits, and others to understand the changing conditions around the pandemic. For example, Google provides a public repository of open-source data related to the global response to the novel coronavirus.[28] This includes epidemiology and health datasets, as well as data on government responses. We also share a privacy-safe version of Google's own mobility data. Google launched the COVID-19 Community Mobility Reports[29] in April of 2020.[30] These reports provide aggregated, anonymized mobility data. They use differential privacy to protect our user data, adding artificial noise to our datasets and enabling high quality results without identifying any individual person.[31] The reports have helped governments around the world understand how social distancing measures are influencing outcomes.

Companies of all sizes utilize code, tools, and knowledge sharing in open source platforms to support their product development. Google is one of the largest contributors to open source code to GitHub, a popular repository for software development tools.[32] In 2020, Googlers made more than 240,000 contributions to tens

[25] With this data, computer vision researchers can train image recognition systems. Waze for cities (https://www.waze.com/en-GB/wazeforcities) datasets are to inform mobility projects and policies, from congestion to event-specific traffic controls. Alphabet's Waymo Open Dataset (https://waymo.com/open/about/) contains high resolution sensor data collected by Waymo self-driving cars to aid the research community in making advancements in machine perception and self-driving technology.
[26] https://trends.google.com/trends/
[27] https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=%22google+trends%22&btnG=
[28] https://cloud.google.com/blog/products/data-analytics/publicly-available-covid-19-data-for-analytics
[29] https://www.google.com/covid19/mobility/
[30] https://blog.google/technology/health/covid-19-community-mobility-reports/
[31] https://support.google.com/covid19-mobility/answer/9825414?hl=en&ref_topic=9822927
[32] https://github.com/google

of thousands of projects on GitHub.[33] Google's open source work includes AI building blocks, data processing tools to process data including TensorFlow Privacy,[34] and models for language understanding and computer vision. All of these contributions make it easier for new and existing companies to develop and bring new products to consumers.

Data alone does not guarantee better products for consumers

Consumers have many choices over where they share their data—and they can and do choose to share the same data with multiple providers. In our experience, data by itself does not guarantee better or more successful products. It is the investment, innovation, and method that matters most.

Cutting edge technology or new ideas allows new companies to succeed, sometimes without any data at all. New entrants such as Zoom, Snapchat, Spotify, or Pinterest have been successful because they provide an innovative product, not because they have access to data from established companies.

Conversely, having more data does not itself guarantee success. Many businesses have had access to data but have had products that struggled to succeed.

Drawing insights from raw or aggregated data, including publicly available data, is where a company can add particular value to consumers and businesses. The process of using information and feedback to improve products and services is not unique to Google or the online world—it takes place across all sectors of our economy, where data helps firms improve and innovate and better serve consumers. It is what a company makes of the data, not how much data they have or use that determines their ability to innovate and succeed, and better serve their customers.

For example, data is just one component of what makes Google Search a useful product for users and businesses. To deliver Search results, Google's systems sort through hundreds of billions of webpages in our Search index to find the most relevant and useful results for our users. The publicly available web is an important data set to make that happen and we also have invested in indexing a wide range of information

---

[33] https://opensource.googleblog.com/2021/08/metrics-spikes-and-uncertainty-open-source-contribution-during-a-global-pandemic.html

[34] https://blog.tensorflow.org/2019/03/introducing-tensorflow-privacy-learning.html

from images, books, research papers, and much more to help our users understand the world's information.[35]

We know we need to keep innovating to find new ways for Search to deliver more useful information faster and more conveniently to consumers. In 2020, we ran over 600,000 experiments that resulted in more than 4,500 improvements to Search.[36] Improving our product involves steps as seemingly simple as a synonym system that allows us to find relevant documents even if they do not contain the exact words a user typed. For example, a user may search for "change laptop brightness" but the manufacturer has written "adjust laptop brightness." Our systems understand the words are related and are able to connect users with the right content. This system took over five years to develop and significantly improves results in over 30% of searches across languages.[37]

Our focus on continually improving our products means that our greatest source of innovation comes from extensive research and development (R&D). Last year alone we spent $27.6 billion on R&D, nearly ten times what we spent in 2009.

Advancements in AI and machine learning increasingly support ways to minimize the use of data while continuing to deliver and improve helpful products people rely on every day. For example, Google has been a leader in federated learning, a machine learning approach that learns from a user's interaction with a given device while keeping all the training data on the device, so that the data does not need to be shared with a server.[38] More recently, Google published research on Entities as Experts AI,[39] that answers text-based questions with less data.

Conclusion

We appreciate the opportunity to share our views on the implications of data for consumers and competition.

---

[35] https://www.google.com/search/howsearchworks/how-search-works/organizing-information/
[36] https://www.google.com/search/howsearchworks/how-search-works/rigorous-testing/
[37] https://www.google.com/search/howsearchworks/how-search-works/ranking-results/
[38] https://ai.googleblog.com/2017/04/federated-learning-collaborative.html
[39] https://venturebeat.com/2020/04/20/googles-entities-as-experts-ai-answers-text-based-questions-with-less-data/

We recognize policymakers and regulators are working to protect privacy while also considering arguments by those who contend that making more data available among competitors would increase competition. These complex discussions involve defining the types of data at issue, as well as identifying the types of services that use data, and the data necessary to make those services useful. We are encouraged to see some privacy and competition regulators conferring more formally to contribute their relevant expertise to the important questions being considered, for example when a privacy practice is being evaluated in an antitrust context. We will continue to engage with policymakers and regulators, as well as other stakeholders, to support thoughtful regulation that encourages innovation and protects consumers. For example, we have long supported federal privacy legislation in the U.S.[40]

We are committed to protecting data through privacy, security, and user control, and to continuing to improve our products in a way that ensures more choice and competition. We look forward to engaging with this Committee on these important issues.

Thank you for the opportunity to discuss our work today.

[40] https://blog.google/competition/#facts