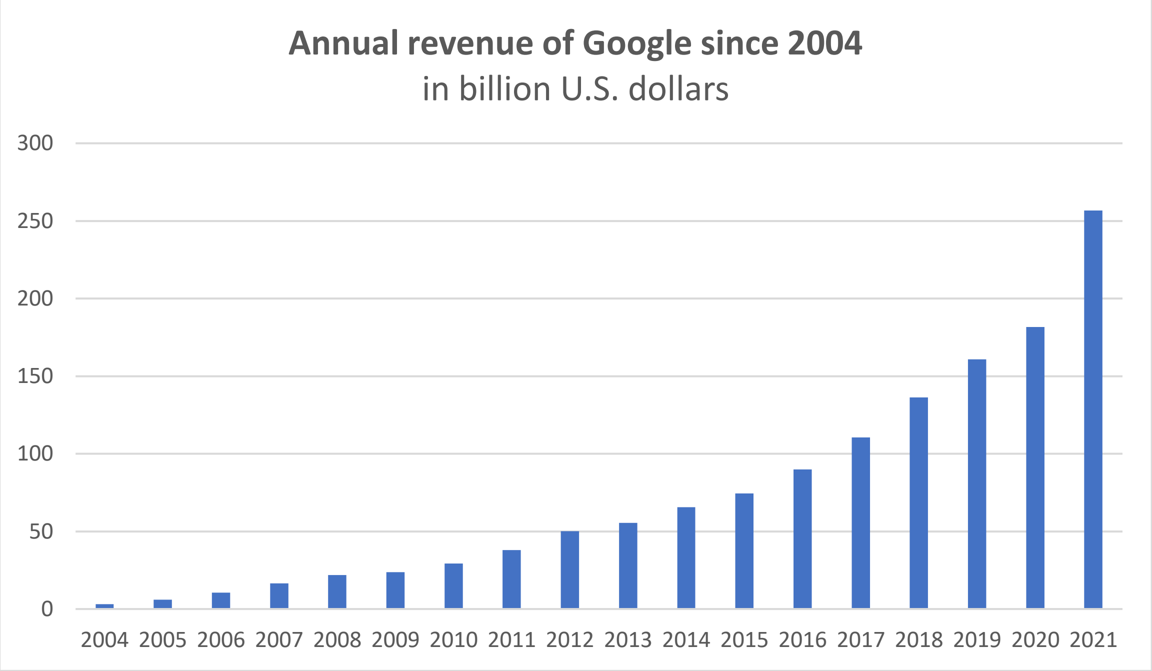
This text is part of the study project »Data Ethics Research Lab« at University of Osnabrück

Why you should opt-out of Google services

The man you see on the flyer is Sundar Pichai. Since 2015, he is the CEO of Google, but he already started his Google career in 2004. Back then, he was in charge of *Google Chrome*, *Google Chrome OS*, and *Google Drive*. Later, he added *Gmail* and *Google Maps* to his portfolio.

Since Pichai started working at Google, its annual revenue has grown relentlessly. While Google's revenue was $3 billion per year in 2004, it had already increased twenty-fold ten years later ($66 billion in 2014). In 2021, it amounted to $256.74 billion per year – which is more than the GDP of Portugal ($252 billion in 2022). Google's revenue is largely based on advertising. In 2021, Google's advertising revenue amounted to $209 billion. The question is: Why do advertisers pay Google so much to place their ads?

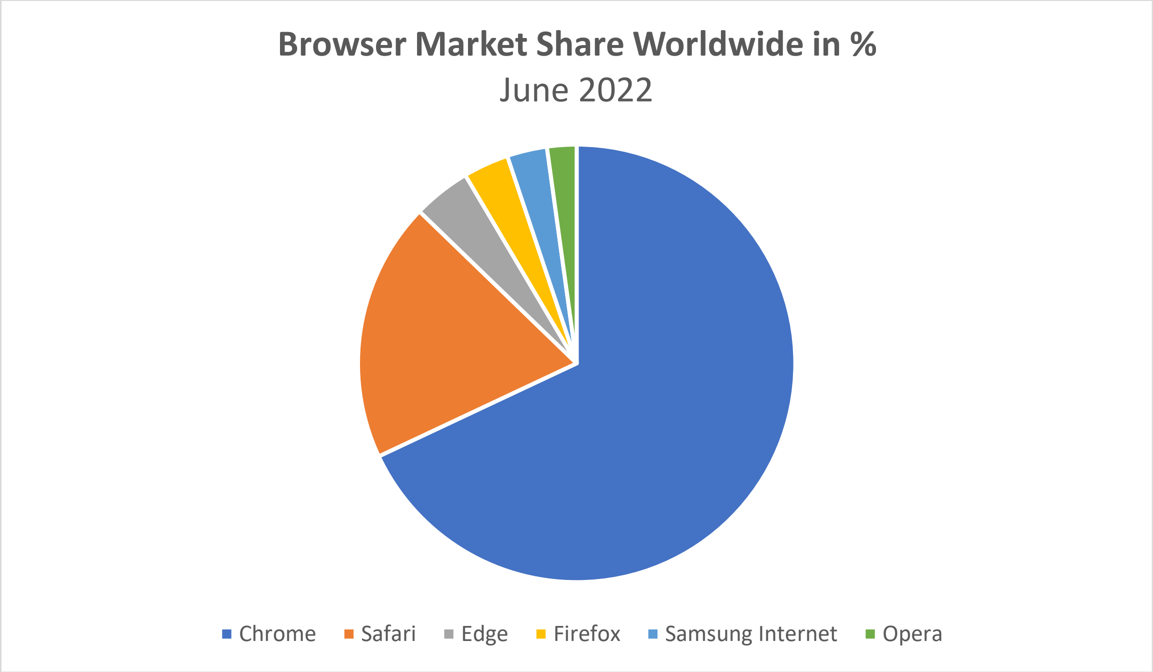


Source: https://www.statista.com/statistics/266206/googles-annual-global-revenue/

Because Google knows everything about its users. All of Google's services are highly developed data collection applications. As soon as you log into any Google product, such as Google Mail, Google Chrome, or Youtube, Google tracks your every move on the Internet:

- **Gmail:** Did you email your friend that you are feeling depressed?
- **Google Chrome:** Have you searched the internet for therapists?
- **Youtube:** Did you watch hours of self-help videos on youtube?

Such user behavior is very likely to be classified as ›depressive‹ by Google's algorithms. A study from 2013 already shows that very precise statements about Facebook users can be made only based on Facebook likes. The study authors were able to predict with 88% accuracy whether users were gay, with 93% accuracy whether the user was a man or a woman, and with 95% accuracy whether the user's ancestry was Caucasian or African American. By using products like Google Chrome, we voluntarily disclose this data about ourselves – and so do two-thirds of all global internet users.



Source: https://gs.statcounter.com/browser-market-share#monthly-202110-202110-bar

How Google Monetizes Your Data

In 2019, tech journalist Geoffrey Fowler ran an experiment: »In a week of Web surfing on my desktop, I discovered 11,189 requests for tracker ‘cookies’ that Chrome would have ushered right onto my computer but were automatically blocked by Firefox.«

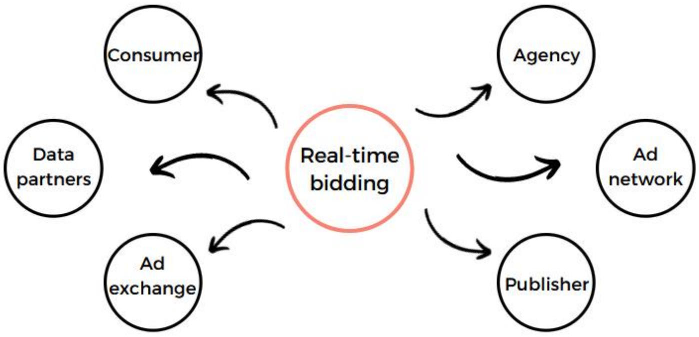
These trackers are like fishing nets. But instead of catching fish, they catch your data. It’s not only Google that uses Cookies, but many websites that want to gather data about their users. 41% of all Internet websites use cookies, primarily to tailor advertising to you.

Google then sells its collected data to advertisers in two different ways:

1. **Google uses your data to create psychological profiles**, e.g., your political attitude, your sexual orientation, your (mental) health, etc. Advertisers can then use Google's algorithms to target groups of people based on these characteristics.

> *When you click on certain pages in Google search, this is feedback for the platform as to whether the search result is relevant to you or not. Similar feedback is gathered when using Google Chrome, Gmail, Youtube, Google Drive, and Google Maps. This feedback is then correlated with other parameters collected about you. And the more data available, the more accurate their psychological profile of you becomes, which is then sold to the highest bidder.*

2. **Google shares data with advertisers directly:** Publishers auction off ad space in their apps or websites, and share your data (geolocation, browsing history, device IDs) with ad tech companies. Advertisers pay a set price for 1,000 people to see their ad via an automated digital auction process. This process is called real-time bidding and Google controls large parts of its ecosystem.



Dangers to the Individual

Data is sold to the highest bidder via data brokers. The buyers are primarily advertisers – but it may also be the Donald Trump or Brexit campaign that want to use the data for ›personalized political advertising‹. Here, political movements advertise with targeted clips on social media, which preferably are only seen by those for whom they have an effect.

The possibility to buy data at any time is not only a danger for election manipulation but also opens the door for hackers and spies, who can use this information to blackmail users. This can be done via so-called ›linkage attacks‹ (also known as ›re-identification› or ›de-anonymization‹ attacks).

In the last 20 years, there has been a rise in re-identification attacks. In 2013, researchers from MIT managed to identify 95% of 1.5 million cellphone users based on their data. The tech journalist Geoffrey Fowler writes that »if you use Android, Chrome sends Google your location every time you conduct a search. (If you turn off location sharing it still sends your coordinates out, just with less accuracy.)« Even coarse datasets provide little anonymity. In 2019, a model from *nature* researchers correctly re-identified 99.98% of Americans in any dataset using 15 demographic attributes.

Linkage attacks

»Linkage attacks work by linking a not-yet-identified dataset (eg. a database of supposedly anonymous medical health records) with some easier-to-obtain auxiliary information on specific individuals (e.g. the day and time that a politician gave birth). The attack is then simply performed by looking for overlapping matches between the common attributes of these two sources of information. Once such a match is found, the direct identifiers can be attributed to the supposedly anonymous data records. In the previously stated example, finding a subject that gave birth at the same date and time as the politician, would then allow to attribute all the other medical records of that subject to the named politician - even though no direct identifiers were contained in the accessed database. Anyone with a basic knowledge of data querying techniques can perform such a ›hack‹, thus it is certainly ›reasonably‹ likely to be performed by a malicious actor.«

Source: https://mostly.ai/blog/synthetic-data-protects-from-ai-based-re-identification-attacks/

Dangers to the Collective

In order to take better care of your privacy, Google Chrome introduced a new API (application programming interface) in 2022: ›Google Topics‹. It works without third-party cookies and only uses *topical interests* based on your recent browser history. These classifiers can be read by humans and list up to several thousand topics. Websites can request Google to get three topics about the user every week, and the respective topics will be deleted after three weeks.

At first glance, this seems to be an improvement in user privacy. However, privacy is also used as an excuse for pushing competitors (that use cookies) out of the market, because now only Google controls all the data streams. Therefore, ›Google Topics‹ is only an improvement in comparison to third-party tracking, but not in comparison to an Internet where behavioral profiling is forbidden.

Predictive Analytics:

»Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques and machine learning.«

Source:
https://www.ibm.com/analytics/predictive-analytics

Besides pushing competitors out of the market, Google Topics also does not solve the collective problems of data collection which arise from predictions of the Google algorithms. So-called ›predictive analytics‹ are based on statistical inferences. For example, users who Google »gay porn« and watch queer documentaries on YouTube are more likely to be classified as gay. Based on these predictions, users receive personalized ads and search suggestions.

Since there is only ever data about what has happened, not what will happen, algorithms learn only from this ›conservative‹ data and predict outcomes based on it. This results in the risk that existing biases will be perpetuated into the future, which, in turn, can lead to stereotypes, determinacy, and self-fulfilling prophecies.

Furthermore, the data that is gathered about *you* is used to make predictions about *other people*. Once the prediction models get sophisticated enough, they don’t necessarily need data about you to predict your personal attributes. Accordingly, predictive analytics does not necessarily violate your individual privacy, but the predictive privacy of arbitrary individuals. So, by seemingly ›solving‹ privacy issues like the re-identification problem, Google prevents its prediction and social sorting methods from entering public discourse Thus, while Google Topics might make individual privacy better, it degrades collective privacy because it obscures the more subliminal goals of Google's data collection.

What Can You Do?

1. **Consider switching your browser**

Mozilla Firefox	Google Chrome
Private Browsing mode	Incognito mode
Blocks third-party tracking cookies by default	–
Blocks cryptomining scripts	–
Blocks social trackers	–

Source: https://www.mozilla.org/en-US/firefox/browsers/compare/chrome/

➔ How to import bookmarks from Chrome to Firefox: https://support.mozilla.org/en-US/kb/import-bookmarks-google-chrome

2. **Consider switching your search engine**

DuckDuckGo	Google Search
Financed through advertising	Financed through advertising
No customized advertisements or search results	Customized advertisements and search results
Does not retain IP address or other user information	Allows third parties to track, store, and sell personal data

Source: https://www.bloggeroutreach.io/blog/duckduckgo-vs-google/

3. **Consider using a VPN**
   - ➔ PC Mag's list of best VPNs in 2022: https://uk.pcmag.com/vpn/138/the-best-vpn-services
   - ➔ But most importantly: Don't surf the web while logged in to Google (or elsewhere).

4. **Increase your privacy settings in Google**
   - ➔ https://www.cloudwards.net/how-to-erase-your-google-history/

More Information On:

a. Data Brokers (YouTube video)
b. Google Chrome (Comic)
c. Google Search (Documentary, only in German)
d. Big Data Ethics (Academic paper)