IARIW – CIGI 2023

IARIW – CIGI 2023

Thursday, November 2 – Friday, November 3

Data Enclaves: Valuing Google's Data Assets

Kean Birch
(York University)
kean@yorku.ca

Paper prepared for the Conference on The Valuation of Data November 2 – November 3, 2023
Session 2: Methodologies of Data Valuation - I
Time: Thursday, November 2, 2023 [11:45AM-1:15PM EST]

** Please do not cite without permission **

Data Enclaves: Valuing Google's Data Assets

KEAN BIRCH
Institute for Technoscience & Society, York University, Canada
kean@yorku.ca

Abstract

Digital data is increasingly understood as a key asset in our digital economies. But how should we value such digital data? It is quite an important question for which no-one currently has a generally accepted answer. Numerous policymakers, regulators, and stakeholders are all trying to work out how to manage the collection, use, and valuation of digital data in order to balance the negative implications of its collection and use – such as privacy loss, data breaches, or declining market competition – with the social and economic benefits that may result – such as improved service delivery, more efficient welfare systems, or better products. A useful way to understand digital data is as an asset. Understanding digital data as an asset requires an interdisciplinary approach because, as an asset, digital data is constituted by a socio-technical configuration of legal rights, technoscientific devices, and policy regimes that create forms of de facto data exclusivity and control through the construction of limitations on data access. Empirical examples of digital data valuation do not reflect ownership and property rights per se, but rather diverse modes of access and use restrictions created through economies of scale, network effects, intellectual property, limited interoperability, contractual arrangements, etc. As a result, digital data is increasing concentrated into 'data enclaves' – rather than data monopolies – controlled by a few, large digital technology firms, especially so-called Big Tech. In this paper, I use Google as a case study of how Big Tech firms treat and value digital data as an asset by analysing how social actors understand, frame, and value the digital data Google collects from users, especially as part of a broad monetization strategy. To do this, I qualitatively analyse an extensive dataset of financial documentary materials produced by and about Google, including: their financial reports (2004-2022); their earnings calls (2005-2022); their SEC filings for mergers (e.g. purchase price allocations); and court cases about the collection and monetization of digital personal data. I use qualitative data analysis software (NVivo and ATLAS.ti) to identify the different modes of access and use restrictions that Google deploys to turn digital data into an asset.

Introduction

Digital data is increasingly understood as a key asset in our digital economies. But how should we value such digital data? It is quite an important question for which no-one currently has a generally accepted answer (Girard 2018; Marciano et al. 2020). Numerous policymakers, regulators, and stakeholders are all trying to work out how to manage the collection, use, and valuation of digital data in order to balance the negative implications of its collection and use –

such as privacy loss, data breaches, or declining market competition – with the social and economic benefits that may result – such as improved service delivery, more efficient welfare systems, or better products. These policy institutions, standards-setters, and stakeholders include sub-national and national governments and their agencies (e.g. UK Treasury, Canadian Competition Bureau; Province of Ontario); national and international statistical offices (e.g. StatCan, System of National Accounts); think tanks (e.g. Centre for International Governance Innovation, Bennett Institute); inter-governmental institutions (e.g. UNCTAD); supra-national institutions (e.g. EU, Eurostat); international institutions (e.g. World Economic Forum, OECD); professional organizations (e.g. CFA, Law Commission); and private business (e.g. Deloitte) (see WEF 2011; Canadian Competition Bureau 2017; UK HM Treasury 2018, 2019; UNCTAD 2018, 2021; CIGI 2019; CFA 2021; OECD 2019a, 2019b, 2022; StatCan 2019; Bennett Institute 2020; Deloitte 2020; EC 2020; European Parliament 2020; Ontario 2021; Law Commission 2021; SNA 2021;). Increasingly, these stakeholders etc. are conceptualizing digital data as an asset, especially in future national accounting standards produced by international policymakers (e.g. System of National Accounts).

Understanding digital data as an asset requires an interdisciplinary approach. Analysing its configuration and valuation as an asset entails drawing on the conceptual and empirical insights of disciplines like science and technology studies (e.g. Beauvisage & Mellet 2020; Birch and Muniesa 2020; Prainsack 2020; Birch et al. 2021), economics and political economy (e.g. Durand & Milberg 2020; Strauss et al. 2021; Rikap 2022), law (e.g. Cohen 2019; Pistor 2019; Drexl 2021; Viljoen et al. 2021), and policy studies (e.g. Haskel & Westlake 2018; Li et al. 2019; Birch et al. 2020; Komljenovic 2022). This interdisciplinary literature stresses the need to think about the techno-economic construction of assets in order to understand them as political-economic objects. As an asset, for example, digital data is constituted by a socio-technical configuration of legal rights, technoscientific devices, and policy regimes that create forms of de facto data exclusivity and control (Drexl 2021), largely through the construction of limitations and restrictions on data access (Zech 2017; Marciano et al. 2020; Birch & Cochrane 2022). As such, actually existing digital data valuation does not reflect the usual forms of ownership and property rights per se, but rather diverse modes of access and use restrictions created through economies of scale, network effects, intellectual property, limited interoperability, contractual arrangements, etc. As a result, digital data has become increasing concentrated into 'data enclaves' (Birch forthcoming) – rather than data monopolies – controlled by a few, large digital technology firms, especially so-called Big Tech (US House of Representatives 2020). This raises a number of implications not only concerning privacy and data protection but also the societal benefits that more open and responsible collection and use of digital data are meant to provide.

In this paper, I use Google as a case study of how Big Tech firms treat and value digital data as an asset by analysing how social actors (e.g. Google itself, financial analysts, investors, etc.) understand, frame, and value the digital data that Google collects from users, especially as part of its advertising monetization strategy. To do this, I qualitatively analyse an extensive dataset of financial documentary materials produced by and about Google, including: their financial reports (2004-2022); their earnings calls (2005-2022); their SEC filings for mergers (e.g. purchase price allocations); and court cases about the collection and monetization of digital

personal data. My starting point was to use qualitative data analysis software (NVivo and ATLAS.ti) in an attempt to identify the different modes of access and use restrictions that Google deploys to turn digital data into an asset. However, in seeking to undertake this analysis, it became evident that there is no easy way to analyse the value of digital data held by Google. Consequently, I end the paper by raising a number of conceptual issues with treating personal data as a political-economic object, whether asset or otherwise.

Digital Personal Data as an Asset

Definitions of Personal Data

Today, when people use the terms like "personal information", "personal data", etc. (henceforth personal data) they generally refer to the "digital" collection, use, and exploitation of specifically "digital" personal data. Recent digital and algorithmic technologies have enabled the massification of data collection, use, and commercial exploitation of personal data, entailing new technological and economic objectives of use (e.g. inferential analytics enabled by 'big' data) and new technological and economic structures of collection (e.g. collection of data on our online and cellular activities) (Birch et al. 2021). Personal data can be categorized in different ways, depending on (1) its type; (2) the main collection method; and (3) its characteristics.

- First, it is possible to identify different types of personal data, reflecting the increasing digital nature of its collection, use, and commercial exploitation. According to the OECD (2013: 8), these include: "User generated content"; "Activity or behavioural data"; "Social data"; "Locational data"; "Demographic data"; and "Identifying data of an official nature" (also see Eben 2018.
- Second, personal data can be characterized by its collection method. Again, the OECD (2013, 2022) is useful here as it differentiates data collection into: "Volunteered" by individuals willingly; "Observed" about individuals; and "Inferred" from the analysis of individuals.
- Finally, personal data can be understood by its characteristics and differentiated between: "Identifiable"; "Anonymous"; and "Pseudonymous" (Edwards 2018).

Personal Data as a Political-economic Object

Academic, business, and policy debates about the commercial uses, exploitation, and value of personal information have been going on for more than two decades (see Spiekermann et al. 2015). Earlier discussions seemed to focus on the notion of personal data markets as markets for privacy, while more recent discussions have expanded into other areas. Generally, it is now commonly accepted that personal data is an important economic resource, asset, and/or commodity across this academic, business, and policy literature (e.g. OECD 2013, 2022; UNCTAD 2021; Coyle and Manley 2022). Recent discussions also tend to emphasize that personal data has certain qualities that make its commercial use and exploitation different from other economic resources, assets, and/or commodities, including:

- Personal data is **non-rivalrous**;
- Nevertheless, personal data can be **excludable**:
- There are **no property rights** to personal data per se;[1]
- Personal data has **emergent properties**, or positive externalities;.
- Personal data is **relational**.[2]

Personal data is an important resource for a range of business models in the digital technology sector, underpinning consumer services like online search; social networking platforms; online advertising, especially with individual targeting; analytical and marketing services for online (and in-person) businesses; and artificial intelligence or algorithmic technology products and services (Beauvisage and Mellet 2020; Eben 2018; Laney 2018; Mattiuzzo 2017; OECD 2013; Spiekermann et al. 2015; Xiong et al. 2022).

The commercial use of personal data is most obvious and prevalent in online advertising (McGuigan 2023), especially with the advent of programmatic advertising. Programmatic advertising took off around 2015 and entails automating the online advertising process and depends upon user data as online advertising has shifted from *context-based* advertising to targeting individual *users* (Srinivasan 2020). Programmatic advertising uses personal data (e.g. user preferences, demographics, location, etc.) to sell online ad inventory to advertisers, meaning that "adtech" companies with better targeting data will attract more advertising transactions, which then potentially reinforce their competitive position by providing them with even more personal data (and then interested advertisers) (ibid.).

Personal data is an important economic resource, asset, and/or commodity for online advertising companies like Alphabet/Google. In 2023, the Bundeskartellamt (2023), Germany's federal competition authority, noted that "Google's business model relies heavily on the processing of user data" and that "Due to its established access to relevant data gathered from a large number of different services, Google enjoys a strategic advantage over other companies". This statement follows an earlier one, in 2021, by the EU's competition authority that the European Commission "already considers data as an asset in merger assessments"; this was in response to a Parliamentary question regarding an investigation of "the way data concerning users is gathered, processed and monetised by Google".[3]

Google as Data Enclave

Adtech and Online Advertising

[1] Some countries and jurisdictions allow property rights for databases, representing a particular arrangement and structuring of data equivalent to copyright (OECD 2022).
[2] Sources for these bullet points include Birch et al. (2021), Cohen (2019), Determann (2018), OECD (2022), and Viljoen (2020).
[3] In 2021, European Commission Executive Vice-President Vestager specifically noted that "The Commission already considers data as an asset in merger assessments", in response to a Parliamentary Question [E-000274/2021(ASW)] about personal data: https://www.europarl.europa.eu/doceo/document/E-9-2021-000274-ASW_EN.html

Online advertising is split between "search" advertising (i.e. matching search keywords with advertising bids) and "display" advertising (i.e. text, image, or video displays on websites, apps, etc.). Online advertising is dependent upon the "adtech" sector ('advertising technology'), which covers a range of intermediaries that sit between advertisers (who buy online ad space) and publishers (who sell online ad space, called "ad inventory") (ACCC 2020; CMA 2020). Online advertising can be traced back to 1994 when advertisers and publishers dealt with each other directly to buy and sell ad inventory. Advertisers could bought ad space on a "cost-per-mille" (CPM) basis, where "mille" represents 1000 impressions or views of a website.

In the second half of the 1990s, adtech intermediaries (e.g. ad servers, ad networks) began to appear as websites proliferated.[4] DoubleClick was an early adtech company, founded in 1995 as an ad network and then acquired by Google in 2008 (Geradin and Katsifis 2019, 2020). Ad networks like DoubleClick acted as intermediaries between advertisers and publishers, aggregating publishers' "remnant" ad inventory (i.e. ad space that publishers could not sell directly) and enabling both advertisers and publishers to outsource and automate the buying and selling of this leftover ad inventory. Gradually the adtech sector expanded as advertisers and publishers sought to optimize their activities and to find more efficient and profitable ways to buy or sell ad inventory. This meant that publishers, through their ad servers, ended up having to work with a range of ad networks to sell all their inventory in what has been called a "waterfall" system: here, publishers would sell their most valuable ad inventory directly, then the rest of their ad inventory would be sold through ad networks and ad exchanges in a gradually declining fee cascade (Geradin and Katsifis 2019, 2020; Bitton and Lewis 2020; Srinivasan 2020). However, the choice of ad network or exchange was often based on historical average performance, which meant that publishers would not always get the best bid for their ad inventory. Google introduced "dynamic allocation" in 2007 as a way to replace these historical average performance metrics with real time bids in their ad exchange (called AdX at the time) (Geradin and Katsifis 2019, 2020; Bitton and Lewis 2020; Srinivasan 2020).

Such "real-time-bidding" (RTB) took off and replaced the waterfall system from the late 2000s onwards, requiring the establishment of industry protocols like the OpenRTB Consortium. RTB became the main way to sell remnant ad inventory and then to sell most ad inventory. RTB enabled the auctioning of a publisher's ad inventory during the milliseconds before a webpage loaded and involves a range of adtech intermediaries, as outlined in Figure 1 below.[5]

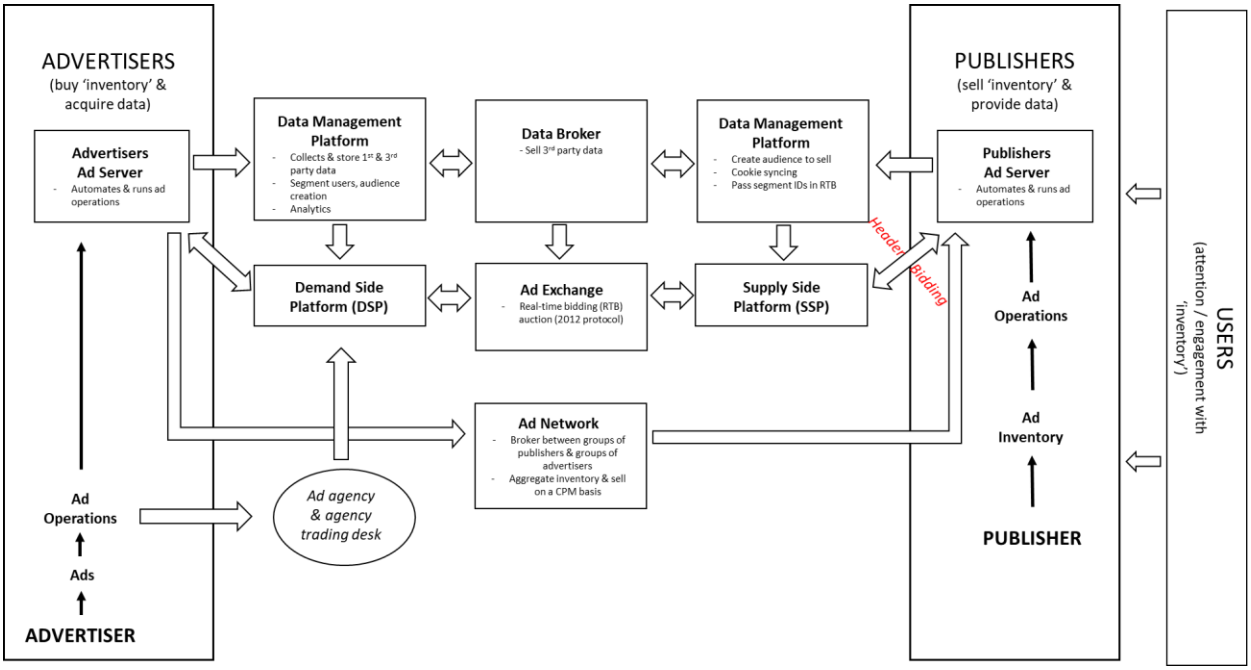
These intermediaries include, from left to right in Figure 1:

- Traditional **advertising agencies** hired by advertisers to manage ad campaigns.
- **Ad servers (advertisers)** to help advertisers manage and automate ad campaigns.

[4] See ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/.
[5] Geradin and Katsifis (2020); also see ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/.

- **Demand-side platforms** (DSP) to buy ad inventory for advertisers (via advertiser ad servers and ad agencies) from ad exchanges or ad networks.
- **Data management platforms** to collect, store, and analyse personal data; they help advertisers (and publishers) create audiences by segmenting users into groups.
- **Data brokers** to collect and trade personal data.
- **Ad exchanges** to operate real-time auctions for ad inventory.
- **Ad networks** to aggregate ad inventory and broker deals between groups of advertisers and groups of publishers.
- **Supply-side platforms** to sell ad inventory for publishers (via publisher ad servers).
- **Ad servers (publishers)** to help publishers manage selling ad inventory.[6]

Figure 1: The Adtech Sector



Source: adapted from ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/

The current state of online advertising represented in Figure 1 is a consequence of the rise of "programmatic advertising" from the late 2000s onwards. Programmatic advertising has automated the online advertising process and is "fuelled by various categories of user data": here, online advertising has moved away from a *context*-based approach (i.e. website content) to an approach based on the characteristics of individual *users* (Geradin and Katsifis 2019: 61). Programmatic advertising works as follows:

- A user visits a website;

[6] See ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/.

- Before the website loads, the user's visit automatically triggers the website publisher's ad server to identify the user (via personal data) and to send its ad inventory to one or more ad exchanges;
- The ad exchanges send requests for bids for that ad inventory to ad buying platforms (which operate on behalf of advertisers looking for users to advertise to);
- The ad buying platforms submit their automated bids to this auction;
- The ad exchanges then pick their winners and return that bid information to the ad server, which picks the highest bid;
- The website finishes loading, showing the ad from the winning bidder (i.e. advertiser).

Since its emergence, programmatic advertising has quickly dominated online advertising, representing 86 percent of online ad revenues in 2022.[7] Programmatic advertising is highly dependent upon the collection and analysis of personal data. In turn, it has driven the expansion of the collection of these various personal data through a range of data collection mechanisms (e.g. cookies, APIs, SDKs, etc.) and architectures (Esteve 2017).

Google and Adtech

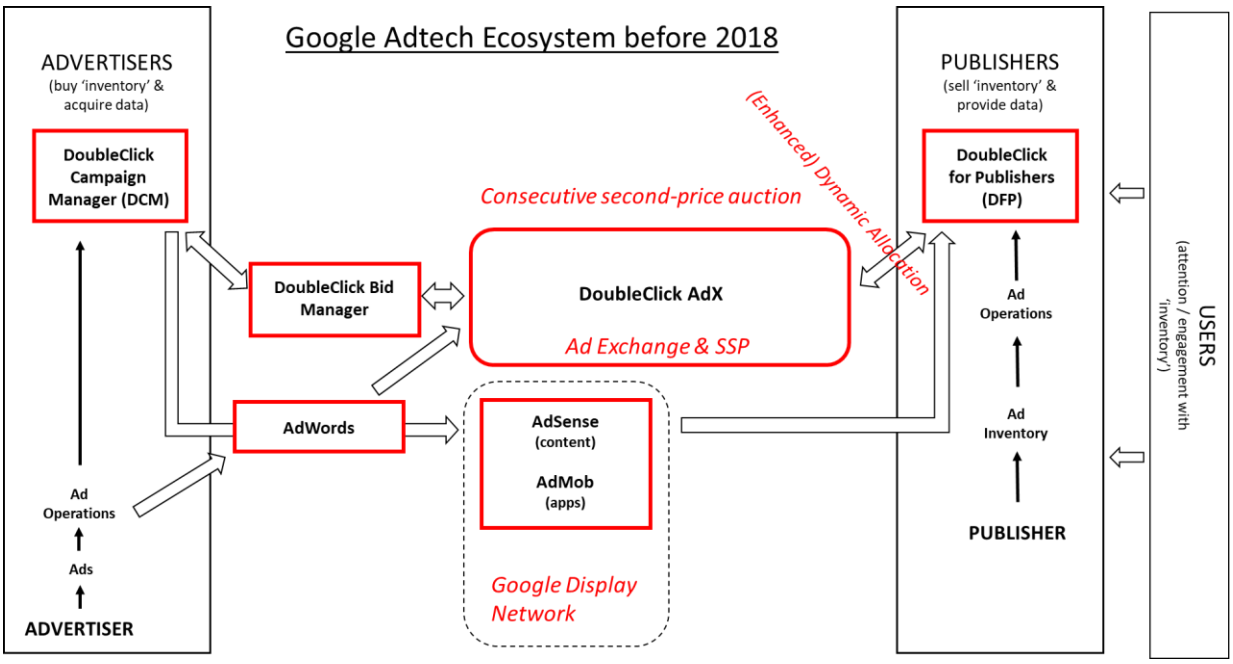
Companies like Google have a competitive advantage in programmatic advertising due to the amount of personal data they collect across their ecosystem. The UK's Competition and Markets Authority (CMA 2020: 155, 280) notes, for example, that Google collects data from over 50 "consumer-facing services", giving it a significant data advantage. Google has been described as "walled gardens" or "enclaves", in that they comprise a range of adtech platforms, products, and services within a single ecosystem.

I illustrate this in Figures 2 and Figure 3 below, differentiating between Google's adtech ecosystem "before 2018" and "after 2018" respectively. Google significantly restructured their adtech ecosystem in 2018 by merging their ad exchange, SSP, and ad server (for publishers) into Google Ad Manager and shifting to a "unified first-price" auction from 2019.[8] Adtech competitors can buy and sell ad inventory within Google's adtech properties, and vice versa, although access to certain Google properties are only available through their own adtech products and services (e.g. YouTube ad inventory through DV360).

Figure 2: Google's Adtech Ecosystem (before 2018)

[7] See the Statista report on "digital advertising" by Chanthadumrongrat (2022).

[8] The type of auction that underpins RTB is an interesting issue in its own right but one that is probably not that useful for this report; suffice to say, Alphabet/Google changed their adtech ecosystem from consecutive second-price auctions prior to 2019 to a unified first-price auction afterwards (Bitton and Lewis 2020; Geradin and Katsifis 2020).



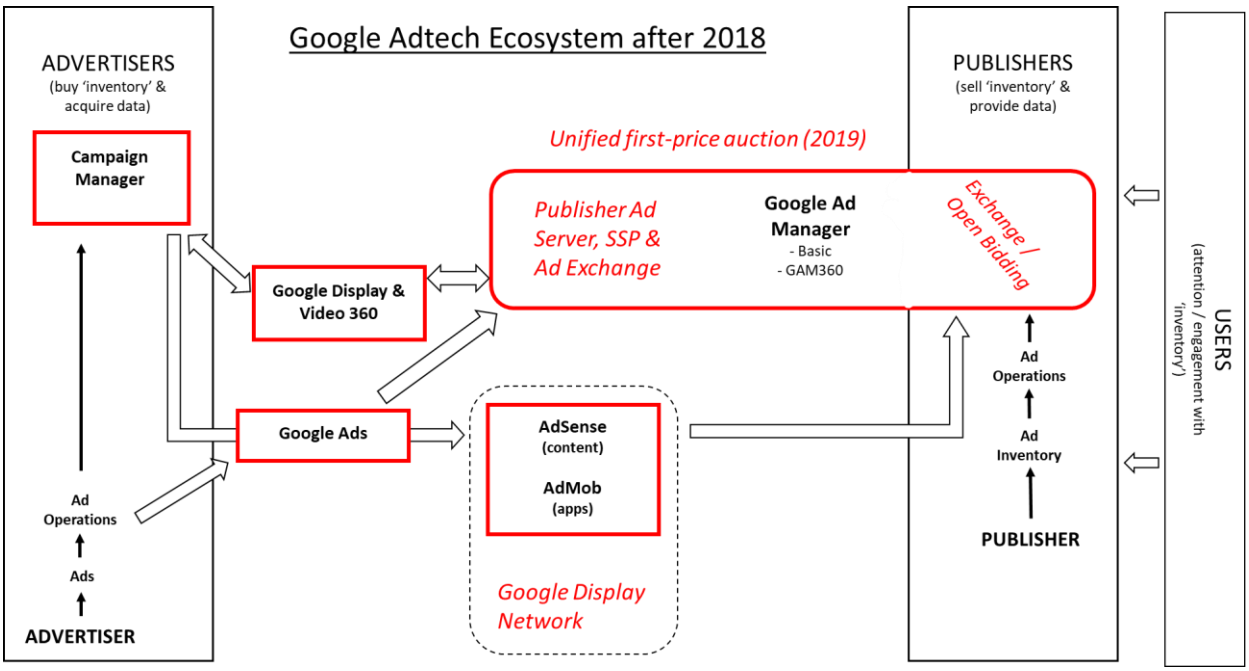
Source: various, including ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/, and Geradin and Katsifis (2019, 2020); Bitton and Lewis (2020); Srinivasan (2020).

Google primarily monetizes personal data through online advertising. On their website, Google's state that "We use data to show ads that are useful to you, whether they are on Google or they are on websites and mobile apps that partner with us. We do not sell your personal information to anyone".[9] Monetizing personal information through advertising has been Google's long-term business model; for example, their 2004 10K Annual Report noted that:

> "Concerns about our collection, use or sharing of personal information or other privacy-related matters, even if unfounded, could damage our reputation and operating results. Recently, several groups have raised privacy concerns in connection with our Gmail free email service which we announced in April 2004 and these concerns have attracted a significant amount of public commentary and attention. The concerns relate principally to *the fact that Gmail uses computers to match advertisements to the content of a user's email message when email messages are viewed using the Gmail service*. Privacy concerns have also arisen with our products that provide improved access to personal information that is already publicly available, but that we have made more readily accessible by the public" (p.54, emphasis added).

Figure 3: Google's Adtech Ecosystem (after 2018)

[9] https://howwemakemoney.withgoogle.com/



Source: various, including ClearCode.cc *Adtech Book*, available online at https://adtechbook.clearcode.cc/, and Geradin and Katsifis (2019, 2020); Bitton and Lewis (2020); Srinivasan (2020).

Google's advertising business model (i.e. monetizing users) has remained relatively consistent throughout their corporate existence. Since its initial public offering in 2004, Google has made most of its revenues from advertising. As their 2022 10K Annual Report states:

> "We generate revenues by delivering relevant, cost-effective online advertising; cloud-based solutions that provide customers with infrastructure and platform services and collaboration tools; sales of other products and services, such as apps and in-app purchases, digital content products, and hardware; and fees received for subscription-based products such as YouTube Premium and YouTube TV" (p.28).

In 2015, around 90 percent of Google's revenues came from advertising, which has slowly declined over time. According to the just mentioned 2022 10K Annual Report, Google's total advertising revenue in 2022 was US$224,473 million (p.59), representing 79.4% of their total revenues and split between:

- "Google Search & other, which includes revenues generated on Google search properties (including revenues from traffic generated by search distribution partners who use Google.com as their default search in browsers, toolbars, etc.), and other Google owned and operated properties like Gmail, Google Maps, and Google Play";
    - 2022 = US$162,450 million revenues
- "YouTube ads, which includes revenues generated on YouTube properties; and"
    - 2022 = US$29,243 million revenues;
- "Google Network, which includes revenues generated on Google Network properties participating in AdMob, AdSense, and Google Ad Manager."

- 2022 = US$32,780 million revenues.

Google generate these advertising revenues by monetizing the "traffic" (i.e. users) attracted to (or generated by) a suite of products and services (defined as "properties" by Google) and partner companies in the 'Google Network'. Consequently, Google generates advertising revenues from its own properties and from other non-own properties within its 'Google Network' (i.e. companies that sign-up to its AdMob, AdSense, and other advertising products and services). Critical to the monetization of traffic is the use of personal data to provide better targeting of users for online advertisers.

In a 2011 Q1 earnings call, Google executives stressed the importance of the "signals" (i.e. personal data) coming from users and/or user traffic (defined as "assets") to the development of their products and services:

> "We do see social as very important. Google uses well over 200 signals in terms of how we think about [Search] ranking today. And when we think about identity and relationships, those are our key signals that can and should be integrated in the experience. So it is important, but it's one of the many that we use. *In terms of assets that apply to that, we do have a very, very large number of users coming to our door every day*. A considerable percentage of them are logged-in users that are using multiple of our products. So there is a large variety of signals that we'll be able to use with user support and users seeing value from it to make the overall experience better" (emphasis added).

A key and ongoing concern of Google since at least its IPO, has been something called "traffic acquisition costs" (TAC) (Scott Morton and Dinielli 2020). According to their 2022 10K Annual Report, TAC consists of:

- "Amounts paid to our distribution partners who make available our search access points and services. Our distribution partners include browser providers, mobile carriers, original equipment manufacturers, and software developers.
- Amounts paid to Google Network partners primarily for ads displayed on their properties" (p.29).

In 2022, TAC amounted to US$48,955 million and is defined as a "cost of revenues", which represents the direct costs of selling a product or service to a customer; in Google's case, the customer is advertisers and the cost of revenues entails the attraction of user traffic and the collection and analysis of their personal data (e.g. to create user profiles). TAC is an important business metric for Google, as it seems to reflect a core resource/asset underlying its monetization strategy through advertising. Their 2006 10K Annual Report emphasizes the possibility that TAC will increase if the company cannot improve monetization:

> "In particular, traffic acquisition costs as a percentage of advertising revenues may increase in the future if we are unable to continue to improve the monetization of traffic

on our web sites and our Google Network members' web sites, particularly with those members to whom we have guaranteed minimum revenue share payments" (p.41).

Reducing the proportion of TAC relative to advertising revenues has been a key goal for Google, which is evident across their annual reports and earnings calls; in 2022, TAC represented around 22 percent of advertising revenues (2022 10K Annual Report, p.34) compared with 32 percent in 2006 (2006 Q2 earnings call). TAC includes "revenue sharing" with mobile carriers and original equipment manufacturers (OEM), especially after the advent of smartphones and their growing importance to online advertising, which becomes most apparent to Google between 2008 and 2012 according to their annual reports and earnings calls.

Google: Personal Data Valuation

Data Valuation Approaches

Although it is possible to identify personal data as an important asset for Google (and other adtech firms), it is difficult to calculate the value of that personal data. To illustrate this, I now want deploy different approaches used to value personal data to analyse Google's data holdings, recognizing that there is no agreed or standard method for valuing personal data.[10] Valuation approaches can be split into at least five categories, each of which reflects a set of theoretical and methodological assumptions about personal data (see Coyle and Manley 2022; Feijoo et al. 2014; Fleckenstein et al. 2023; Girard et al. 2021; HM Treasury 2018; Savona 2019). The valuation methods can be split into:

- Business or organizational approaches;
- Market approaches;
- User or data subject approaches;
- Social benefits approaches; and
- Dimensional approaches.

Here, I focus on the first three valuation approaches, each of which can be further disaggregated by different methodological choice; the other two approaches are less relevant to my analysis. I outline the basis of these three approaches and then subsequently apply them to analyse Google's data holdings.

- **Business or organizational** approaches to calculate personal data value from business performance:

[10] See literature on data valuation, including Beauvisage and Mellet (2020), Birch et al. (2021), Coyle and Manley (2022), Eben (2018), Feijoo et al. (2014), Fleckenstein et al. (2023), Girard et al. (2021), HM Treasury (2018), Laney (2018), Marciano et al. (2020), Meredith (2022), Metzger (2017), OECD (2022), Savona (2019), and Xiong et al. (2022).

    - Recorded **assets** of a business (e.g. intangible assets, goodwill): this is not always possible to do with personal data, however, as personal data is not (usually) recorded on company balance sheets.
    - **Market capitalization** (or private valuation) of a business, reflecting the notion that market sentiment (i.e. what investors are willing to pay) is the best way to judge the value of personal data: however, sentiment is subjective and can oscillate.
    - **Revenues** of a business: methods like ARPU ("average revenue per user") have been used to assess the value of personal data; however, they require precise information on revenues accorded to specific personal information.
- … or from business costs:
    - Cost of **producing or replacing** personal information: this reflects the idea that the value of personal information can be calculated from the historical cost to produce it or to replace it if it is lost or destroyed. There are limits to this approach, however, since it reflects the "lower-bound" of valuation (because it is 'historical' and is difficult to measure current market 'fair value') and is often based on aggregate calculations, rather than calculations of specific datasets.
    - Cost of **damages**: this reflects the regulatory or legal fees that companies would be required to pay when they experience a data breach.
    - A version of **relief of royalty**: this reflects a calculation of the costs that a business *would* incur if it had to license data in externally instead of producing it itself; this could be calculated from the cost of personal information in data markets, or a calculation of the cost of production.
    - **Cost of revenues:** this reflects the costs incurred in generating revenues, including investment in the collection of personal information.
- **Market** approaches to calculate personal data value from market transactions:
    - **Sale** of personal data in 'data markets' (e.g. data brokers): personal data may be more valuable than the direct price in such a data market (e.g. by combining new datasets with existing datasets) while data market pricing reflects business-to-business (B2B) transactions, but not user-to-business transactions (U2B) which are more relevant for personal data.
    - I**llegal sale** of personal data on the 'black market': this might include the sale of information like credit card details.
- **User or data subject** approaches to calculate personal data value from preferences of individual users (for privacy) through surveys or experiments (stated preferences):
    - A user's **willingness to pay** (WTP) to protect their personal information against disclosure: this can be calculated through surveys of users, asking them how much they would pay for their privacy.
    - A user's **willingness to accept** (WTA) payment for personal information: this can be calculated through surveys again, reflecting what a user would accept to sell their personal information; this sale price tends to be higher than WTP.

Valuing Google's Data

Here, I apply the three valuation approaches introduced above to Google. Starting with **market** and **user** approaches, which provide limited means to calculate the value of Google's data holdings. Google clearly state that they do not sell personal data, including having a "security and privacy principle" on their website to "Never sell our users' personal information to anyone".[11] This means that **market** approaches are not useful for understanding the value of the company's personal data. Similarly, it would be possible to survey Google's users to work out their **willingness to pay** (WTP) for privacy and **willingness to accept** (WTA) payment for their personal data, but this suffers from the fact that any valuation would reflect a stated preference rather than actual user behaviour.

Business or organizational approaches seem more useful, but not all of them. It is evident that Google does not record personal data as an **asset** on their balance sheet, or frame it as an asset in their annual reports and earnings calls. Google's **market capitalization** could be a better proxy to understand personal data value, but market capitalization reflects investor expectations about a company (and its assets) and does not necessarily help the analysis. In particular, the market capitalization of Google has risen and fallen quite significantly over the past few years, which is unlikely to reflect similar rise and declines in personal data value in the same time period.

Like other digital technology companies, Google's **revenues** can be used to calculate the value of personal data with metrics like ARPU ("average revenue per user"). Two relatively old studies identify an aggregate ARPU for Google: one puts ARPU at US$40 per user (in 2012) and another puts ARPU at US$59 per user (in 2017).[12] Neither is very clear about how this amount is calculated, however. It is easier to identify the ARPU for specific Google products. There is one mention of an ARPU in Google's own annual reports and earnings calls (2004-2022), suggesting the ARPU for Google Play is around US$7-8 (in 2019).[13] It is also possible to calculate the ARPU of products like YouTube, which can be calculated at US$11.63 per user (in 2022) by dividing YouTube's ad revenues (US$29,243 million) by the number of users (2,514 million). Despite the two aggregate studies mentioned above, it is difficult to calculate an aggregate ARPU for Google without better access to specific (and confidential) data on business segment revenues and user numbers. For example, it might be possible to calculate a ARPU for "Google Search & Other" (in 2021) by dividing revenues of US$148.95 billion by the total 2,852 billion searches in the same year. Here, each search would generate around US$0.05 in revenue. A major caveat is that such a calculation does not take into account what constitutes the "Other" in these revenues.

Google's **costs** seem to be a better proxy for calculating personal data value, although which business costs matters. **Production costs** and **relief of royalty** are difficult to calculate because it is difficult to disaggregate tangible and intangible costs of personal data collection and monetization (e.g. data centres, intellectual property). An alternative would be to use **cost of**

[11] https://safety.google/intl/en/principles/
[12] Becceril (2018); and https://www.enisa.europa.eu/publications/info-notes/the-value-of-personal-online-data
[13] 2019 Q4 earnings call.

revenues (specified in Google's annual financial reports) because this identifies costs incurred in generating specific revenues from advertising, Google's primary business activity. The empirical materials I analyse in the paper highlight the importance of "traffic acquisition costs" (TAC) in this regard. TAC represents the costs of attracting users (and their personal data) to Google's adtech ecosystem in order, then, to monetize users by selling access to them to advertisers.

It might be possible to calculate the aggregate value of personal data by using the TAC rate that Google pays each year. TAC reflects the value Google itself is willing to pay for attracting users *and* their personal data to the company's properties in order to monetize them through advertising. To do this properly, however, would require more disaggregated information on user numbers and revenues generated from Google's different properties (e.g. Search, Gmail, Play, Maps, etc.).

It might also be possible to calculate personal data value using the specific payments Googles makes to other companies for access to their users; a key example is the TAC payments Google makes to Apple to set Google Search as a default on iOS devices. (1) According to court documents, Google paid Apple US$1 billion in 2014 to set Google Search as iOS default.[14] Apple sold 192.7 million smartphone units that year,[15] so Google paid US$5.18 per *new* iPhone user. (2) Using Apple's *cumulative* smartphone unit sales, which were 442 million by 2014, mean that Google paid US$3.35 per iPhone user. (3) According to a 2020 US Congressional investigation into competition in digital markets, "Apple also reportedly made $9 billion in 2018 and $12 billion in 2019 to set Google as the default search engine on the Safari browser" (US House of Representative 2020: 345). Apple's cumulative smartphone unit sales were 888 million and 948 million in 2018 and 2019 respectively, meaning that Alphabet/Google paid US$10.13 and US$12.65 respectively per iPhone user.

More generally, assuming that Google's TAC rate, which was around 23 percent of advertising revenues in 2014 and 22 percent in 2018 and 2019, reflects the cost of revenues for *all* users and their personal data to Google, then this would mean that Google derives the following value from each iPhone user and their personal data (cumulative):

- 2014: US$14.56
- 2018: US$46.04
- 2019: US$57.50

While this could represent the value of personal data to Google, there are caveats to these calculations: (1) they only reflect iPhone devices, not all Apple devices; (2) it is probably better to use cumulative users in the calculation, since Google benefits from access to all users; and (3)

[14] https://www.theverge.com/2016/1/22/10813558/google-apple-1-billion-default-ios-search
[15] https://www.statista.com/study/11643/iphone-statista-dossier/ & https://www.statista.com/statistics/1337745/active-apple-iphone-units-worldwide/

total value to Google reflects total advertising revenues from those users and their personal data.

As can be seen from this attempt to work out the value of personal data held by Google, there is a significant range in valuations across different approaches.

Conclusions & Going Forward: How to Value Personal Data?

Despite the possibility of calculating the personal data value of Google's data holding using a number of valuation approaches, the analysis in this paper shows that these valuation approaches are always (and perhaps necessarily) partial. In particular, the analysis is based on using proxies and estimates triangulating across a range of empirical sources (which are often highly ambiguous or partial). It is notable that Google's annual financial reports and earnings calls rarely contain information about personal data and its value, presumably because of the focus Google has on monetizing users rather than data (Birch et al. 2021); it is also likely that Google's own approach (alongside interested social actors like investors and financial analysts) to personal data is framed by assumptions about what personal data is and what makes it important or not as a political-economic object (see Stucke 2022).

Consequently, it is worth considering other empirical methods that could be deployed to get at the value of personal data (when conceptualizing them as assets) as understood by social actors themselves: 'vernacular value'. I initially wanted to get access to the purchase price allocations / agreements (PPA) of mergers and acquisitions by Google, since these should reflect the value allocated to different assets by Google during the M&A process. PPAs are filed with the USA's Internal Revenue Service (IRS) as they can have significant impacts on corporate tax rates and tax planning for companies like Google (Parsons 2022).[16] A PPA consists of a document allocating the value of the assets of a merged or acquired company, divided into seven different asset classes. I would expect that the value of digital data would be allocated to one of these asset classes as part of the M&A negotiation (probably Class VI). However, while Google's annual reports do discuss "purchase price" of acquisitions, there are few details about the allocation across different asset classes and the IRS documents themselves are not public documents.[17]

While my attempt to access PPAs was unsuccessful, it did start me thinking about how to empirically investigate digital data value. To start, it is useful to conceptualize valuation as a political-economic practice that social actors undertake at different moments in time since assets have a particular temporality to them. I can think of at least six *moments of valuation* that could provide useful empirical sites for further investigation and analysis of data value, including:

[16] https://www.irs.gov/instructions/i8594

[17] I sought help from business school librarians who were unable to find a way to access them.

- **Private investment:** during negotiations over startup investment, companies and investors have to agree to the value of their asset base, which includes digital data.
- **Initial public offering:** companies publish prospectuses of their businesses in which they justify their initial market capitalization, including specifying their asset base.
- **M&A:** as mentioned above, companies need to publish purchase price allocations to justify what they pay for another company.
- **Collateral:** companies use their assets to borrow money; there was an example of US airlines using their customer data to underpin their borrowing during Covid. United Airlines and American Airlines used the value of their customer data as collateral to secure billion-dollar loans; in fact, these data were more valuable than the stock market valuation of these corporations themselves.[18]
- **Compensation (data breach):** companies have to pay for harms incurred during data breaches, so there should be some legal characterization of the value of personal data in such decisions.[19]
- **Bankruptcy:** the asset bases of bankrupt companies are assessed, so there should be details about the value assigned to data holdings in these cases.

An empirical examination of these moments of valuation could provide important an important contribution to ongoing debates about personal data value by getting at the valuation judgements of the social actors themselves, rather than relying upon valuation models or experiments.

References

1. ACCC (2020) *Digital advertising services inquiry*, Canberra: Australian Competition and Consumer Commission, available at: https://www.accc.gov.au/publications/digital-advertising-services-inquiry-final-report
2. Beauvisage T, Mellet K (2020) Datassets: Assetizing and marketizing personal data, in K. Birch and F. Muniesa (eds) *Assetization*, Cambridge MA: MIT Press, pp.75–95.
3. Becerril, A. (2018) The value of our personal data in the Big Data and the Internet of all Things Era, *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal* 7(2): 71-80.
4. Birch, K., Cochrane, D. and Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech, *Big Data & Society* 8(1): https://doi.org/10.1177/20539517211017308
5. Bitton, D. and Lewis, S. (2020) *Clearing up misconceptions about Google's ad tech business*, available at: https://www.competitionpolicyinternational.com/clearing-up-misconceptions-about-googles-ad-tech-business/

[18] United Airlines and American Airlines used their customer data as collateral to secure billion-dollar loans; Laney, D. (2020) Your company's data may be worth more than your company, *Forbes* (22 Jul).

[19] To data, US court cases involving harms from personal data collection have generally failed on the basis that plaintiff's lack standing because no harm has incurred (because they can still sell their data themselves, even if others have already collected it).

6. Boissay, F., Ehlers, T., Gambacorta, L. and Song Shin, H. (2021) *Big techs in finance: on the new nexus between data privacy and competition* [BIS Working Papers No. 970], Basel: Bank for International Settlements.
7. Bundeskartellamt (2023) Press release: Statement of objections issued against Google's data processing terms, Bonn: Bundeskartellamt, available at: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2023/11_01_2023_Google_Data_Processing_Terms.html
8. Chanthadumrongrat, K. (2022) *Digital Advertising – Market Data Analysis & Forecast*, Hamburg: Statista.
9. Chopra, R. (2019) Written Testimony of FTC Commissioner Rohit Chopra, Before the U.S. House of Representatives Committee on the Judiciary Subcommittee on Antitrust, Commercial, and Administrative Law, available at: https://www.ftc.gov/system/files/documents/public_statements/1549812/chopra_-_testimony_at_hearing_on_online_platforms_and_market_power_part_3_10-18-19.pdf
10. CMA (2020) *Online platforms and digital advertising market study*, UK Competitions & Markets Authority, available at: https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study
11. Cohen, J. (2019) *Between Truth and Power*, Oxford: Oxford University Press.
12. Coyle, D. and Diepeveen, S. (2021) *Creating and governing social value from data*, available at SSRN: https://ssrn.com/abstract=3973034 or http://dx.doi.org/10.2139/ssrn.3973034
13. Coyle, D. and Manley, A. (2022) *What is the Value of Data? A review of empirical methods*, Bennett Institute for Public Policy, University of Cambridge.
14. Determann, L. (2018) No one owns data, *Hastings Law Journal* 70: 1-43.
15. Eben, M. (2018) Market definition and free online services: The prospect of personal data as price, *I/A: A Journal of Law and Policy for the Information Society*, 14(2): 227-281.
16. Feijóo, C., Gómez-Barroso, J.L. and Voigt, P. (2014) Exploring the economic value of personal information from firms' financial statements, *International Journal of Information Management* 34(2): 248-256.
17. Fleckenstein, M., Obaidi, A. and Tryfona, N. (2023) A Review of Data Valuation Approaches and Building and Scoring a Data Valuation Model, *Harvard Data Science Review* 5(1): https://doi.org/10.1162/99608f92.c18db966
18. Geradin, D. and Katsifis, D. (2019) An EU competition law analysis of online display advertising in the programmatic age, *European Competition Journal* 15(1): 55-96.
19. Geradin, D. and Katsifis, D. (2020) "Trust me, I'm fair": analysing Google's latest practices in ad tech from the perspective of EU competition law, *European Competition Journal* 16(1): 11-54.
20. Girard, M., Lionais, M. and McLean, R. (2021) *What Is Your Data Worth? Insights for CPAs*, Toronto: Chartered Professional Accountants Canada.
21. Godel, M., Litchfield, A. and Mantovani, I. (2012) The value of personal information evidence from empirical economic studies, *Digiworld Economic Journal* 88(4): 41-60.

22. Guay, R. and Birch, K. (2022). A comparative analysis of data governance: Socio-technical imaginaries of digital personal data in the USA and EU (2008–2016), *Big Data & Society 9*(2): https://doi.org/10.1177/20539517221112925
23. HM Treasury (2018) *The economic value of data: discussion paper*, London: HM Treasury.
24. Kak, A. West, S.M. (2023) *AI Now 2023 Landscape: Confronting Tech Power*, AI Now Institute, available at: https://ainowinstitute.org/2023-landscape
25. Laney, D. (2018) *Infonomics*, New York: Bibliomotion.
26. Lassila, E. (2022) "Free"-to-play game: Governing the everyday life of digital popular culture, *Critical Perspectives on Accounting* 87: 102434.
27. Marciano, A., Nicita, A. and Ramello, G.B. (2020) Big data and big techs: Understanding the value of information in platform capitalism, *European Journal of Law and Economics* 50: 345-358.
28. Mattiuzzo, M. (2017) Business Models and Big Data: How Google Users Your Personal Information, available at: https://itsrio.org/wp-content/uploads/2017/03/Marcela-Mattiuzzo-V-REVISADO.pdf
29. McGuigan, L. (2023) *Selling the American People*, Cambridge MA: MIT Press.
30. Metzger, A. (2017) Data as Counter-Performance: What Rights and Duties do Parties Have?, *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 8(1): 2-8.
31. Moro-Visconti, R. (2022) *The Valuation of Digital Intangibles: Technology, Marketing, and the Metaverse* (2nd edition), London: Palgrave Macmillan.
32. Nissenbaum H (2017) Deregulating collection: Must privacy give way to use regulation?, available at SSRN: https://ssrn.com/abstract=3092282 or https://doi.org/10.2139/ssrn.3092282
33. OECD (2013) *Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value*, Paris: Organisation for Economic Co-operation and Development.
34. OECD (2022) *Going Digital Toolkit Note: Measuring the economic value of data*, Paris: Organisation for Economic Co-operation and Development.
35. Pasquale, F. (2015) *The Black Box Society*, Cambridge MA: Harvard University Press.
36. Rocher, L., Hendrickx, J. and de Montjoye, Y. (2019) Estimating the success of re-identifications in incomplete datasets using generative models, *Nature Communications* 10(1): 3069.
37. Roosendaal, A., van Lieshout, M. and van Veenstra, A.F. (2014) *Personal Data Markets*, TNO Report (R11390): TNO Innovation for Life.
38. Savona, M. (2019) *The Value of Data: Towards a Framework to Redistribute It* (October 2019), SPRU Workings Paper Series 21, available at SSRN: https://ssrn.com/abstract=3476668 or http://dx.doi.org/10.2139/ssrn.3476668
39. Scassa, T. (2018) *Data Ownership* [CIGI Papers No.187], Waterloo: Centre for International Governance Innovation.
40. Scott Morton, F. and Dinielli, D. (2020) *Roadmap for a Digital Advertising Monopolization Case Against Google*, Omidyar Network, available at:

https://publicknowledge.org/policy/roadmap-for-a-digital-advertising-monopolization-case-against-google/
41. Spiekermann, S. and Korunovska, J. (2017) Towards a value theory for personal data, *Journal of Information Technology*, 32: 62-84.
42. Srinivasan, D. (2020) Why Google Dominates Advertising Markets: Competition Policy Should Lean on the Principles of Financial Market Regulation, *Stanford Technology Law Review* 24(1): 55-175.
43. Strauss, I., O'Reilly, T., Mazzucato, M. and Ryan-Collins, J. (2021) *Crouching tiger, hidden dragons: How 10-K disclosure rules help Big Tech conceal market power and expand platform dominance* [Policy Report No.2021/04], London: UCL Institute for Innovation and Public Purpose.
44. Stucke, M. (2022) *Breaking Away*, Oxford: Oxford University Press.
45. Sweeney, M. (2022) *Understanding the Complicated World of Advertising Technology (AdTech) & Programmatic Advertising*, Clearcode, available at: https://clearcode.cc/blog/understanding-advertising-technology/
46. UNCTAD (2021) *Digital Economy Report 2021*, New York: UN Conference on Trade and Development.
47. US House of Representatives (2020) *Investigation of Competition in Digital Market*, Washington DC: House of Representatives.
48. Viljoen S (2020) Democratic data: A relational theory for data governance, *Yale Law Review*, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562
49. WEF (2011) *Personal Data: The Emergence of a New Asset Class*. Geneva: World Economic Forum.
50. Winseck, D. (2022a) *Growth and Upheaval in the Network Media Economy, 1984-2021*. Global Media and Internet Concentration Project, Carleton University, https://doi.org/10.22215/gmicp/2022.03.
51. Winseck, D. (2022b) Media and Internet Concentration in Canada, 1984-2021. Global Media and Internet Concentration Project, Carleton University, https://doi.org/10.22215/gmicp/2022.02.

Secondary Sources

1. Google Annual Report 2004 (10K)
2. Google Annual Report 2005 (10K)
3. Google Annual Report 2006 (10K)
4. Google Annual Report 2007 (10K)
5. Google Annual Report 2008 (10K)
6. Google Annual Report 2009 (10K)
7. Google Annual Report 2010 (10K)
8. Google Annual Report 2011 (10K)
9. Google Annual Report 2012(10K)
10. Google Annual Report 2013 (10K)
11. Google Annual Report 2014 (10K)
12. Google/Alphabet Annual Report 2015 (10K)

13. Alphabet Annual Report 2016 (10K)
14. Alphabet Annual Report 2017 (10K)
15. Alphabet Annual Report 2018 (10K)
16. Alphabet Annual Report 2019 (10K)
17. Alphabet Annual Report 2020 (10K)
18. Alphabet Annual Report 2021 (10K)
19. Alphabet Annual Report 2022 (10K)
20. Google Earnings Call 2005: Presentation and Q&A (Q4)
21. Google Earnings Call 2006: Presentation and Q&A (Q1-Q4)
22. Google Earnings Call 2007: Presentation and Q&A (Q1-Q4)
23. Google Earnings Call 2008: Presentation and Q&A (Q1-Q4)
24. Google Earnings Call 2009: Presentation and Q&A (Q1-Q4)
25. Google Earnings Call 2010: Presentation and Q&A (Q1-Q4)
26. Google Earnings Call 2011: Presentation and Q&A (Q1-Q4)
27. Google Earnings Call 2012: Presentation and Q&A (Q1-Q4)
28. Google Earnings Call 2013: Presentation and Q&A (Q1-Q4)
29. Google Earnings Call 2014: Presentation and Q&A (Q1-Q4)
30. Alphabet/Google Earnings Call 2015: Presentation and Q&A (Q1-Q4)
31. Alphabet Earnings Call 2016: Presentation and Q&A (Q1-Q4)
32. Alphabet Earnings Call 2017: Presentation and Q&A (Q1-Q4)
33. Alphabet Earnings Call 2018: Presentation and Q&A (Q1-Q4)
34. Alphabet Earnings Call 2019: Presentation and Q&A (Q1-Q4)
35. Alphabet Earnings Call 2020: Presentation and Q&A (Q1-Q4)
36. Alphabet Earnings Call 2021: Presentation and Q&A (Q1-Q4)
37. Alphabet Earnings Call 2022: Presentation and Q&A (Q1)