INTERNET POLICY REVIEW
Journal on internet regulation

Volume 11 **Issue** 3

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

The untamed and discreet role of data brokers in surveillance capitalism: a transnational and interdisciplinary overview

Urbano Reviglio *University of Milan*

DOI: https://doi.org/10.14763/2022.3.1670

Published: 4 August 2022
Received: 6 September 2021 **Accepted:** 21 December 2021

Funding: Research for this article has been partially funded by the project "Public Perception of Algorithms in Society: Accounting for the Algorithmic Public Opinion – ALGOCOUNT" (2020-2022), funded by Fondazione Cariplo.
Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Citation: Reviglio, U. (2022). The untamed and discreet role of data brokers in surveillance capitalism: a transnational and interdisciplinary overview. *Internet Policy Review*, *11*(3). https://doi.org/10.14763/2022.3.1670

Keywords: Data brokers, Privacy, Data protection, Surveillance, Datafication

Abstract: Data brokers have a significant role in data markets and, more broadly, in surveillance capitalism. Due to increasingly sophisticated techniques, data brokers allow for pervasive datafication. This not only seriously threatens privacy, but also national security and the necessary trust for data markets to function properly. The data broker industry, however, is an under-researched and under-regulated subject. Thus, this article provides an up-to-date critical literature review, highlighting innovative policy proposals and elaborating further research questions. Overall, apart from strengthening privacy protection, the article makes a case for further research on data brokers and a more inclusive international discussion that may eventually lead to a new social contract for data that is focused, above all, on data standardisation, economic incentives, data brokers' legal definitions, and the creation of an oversight authority.

Introduction

Data brokers have a significant role in the data economy and, more broadly, in "surveillance capitalism". Yet, their political role is insufficiently discussed, whereas the industry benefits from a lack of regulation. Even Shoshana Zuboff (2019), who convincingly explains the logic that dominates capitalism in the digital age in her influential essay *The Age of Surveillance Capitalism,* mentions data brokers only twice in the text. And yet, the role of data brokers in the broader data ecosystem is significant, if not essential. They are pivotal actors of what Zuboff calls "markets for future behaviour" (2019). Their role indeed deserves further attention. Thousands of data brokers exist worldwide who are able to infer thousands of individual features from billions of consumers in order to predict their behaviour. The company Oracle, for example, claims to have data on more than two billion people globally and is able to surmise more than 30.000 attributes (i.e. behavioural insights) for each individual (Christl, 2017). These profiles can be exchanged with third parties (e.g. big tech) and other data brokers, and eventually, kept virtually forever. In this sense, the role of data brokers is rather opaque.

Relatively little research has been undertaken to understand the political role of this subject. Still, data brokers are part of a lucrative industry that is believed to generate at least more than 200bn USD in revenue yearly (though the actual amount is difficult to approximate due to their opacity) (Lazarus, 2019). They essentially sell information about individuals to private individuals, as well as corporate and governmental actors. Given the industry's size and that their business is to know everyone's business, data brokers are a fundamental actor of surveillance capitalism. The services they offer are increasingly important for countless decisions that concern human rights. This trend will only continue. Yet, consumers are largely – if not completely – unaware of how data brokers consolidate, aggregate, analyse, and sell their data. This situation has resulted in two main concerns: data brokers invade privacy and consumers are exposed to unwarranted and unexpected discrimination. However, in this article I argue that data brokers compel further concerns, including geopolitical stability and trust in data markets (see Figure 1).

With future developments in data collection (e.g. Internet of Things[1]) and data analysis, the role of data brokers is likely to increase. It is necessary, then, to better

1. The Internet of Things (IoT) refers to the interconnection, via the internet, of computing devices embedded in everyday objects, enabling them to send and receive data. This is expected to lead to an "ambient intelligence" in which automatic smart online and offline environments and devices interact with each other, making an unprecedented number of decisions for us to cater to our inferred preferences, representing a new paradigm in the construction of knowledge (Hildebrandt & Koops, 2010).

understand, question and problematise their role. Academic literature on data brokers is indeed scarce. It is mostly based on journalistic investigations (Smith, 1997; Anthes, 2015; Kroft, 2014; Christl, 2017; Leetaru, 2018; Ng & Varner, 2021) and economic or legal analysis (Hoofnagle, 2003; Tsesis, 2014; Muralidhar & Palk, 2018; Rostow, 2017; Yeh, 2018; Sherman, 2021a). Thus, this article offers an updated multidisciplinary analysis on this subject, combining critical theories with political data economy, regulation and innovative policy proposals. The objective is twofold; on the one hand, I introduce the data broker industry and discuss the persistent policy vacuum. On the other hand, I critically problematise the role of data brokers in surveillance capitalism and provide a systemic and transnational overview of promising policy horizons.

In doing so, the article first provides a comprehensive description of what data brokers are, exploring their techniques for data extraction, as well as the services offered. Second, it introduces the legal landscape and the policy vacuum that currently exists, with a particular focus on the United States (US) and the European Union (EU). Third, it discusses the political role of data brokers in the context of surveillance capitalism. And finally, I discuss policies and research directions to reframe the debate, especially in the light of forthcoming regulations, such as the European Data Governance Act.

Section 1: Data brokers: who, how and why

No authoritative or comprehensive definition of “data brokers” yet exists. This is mainly due to the fact that data brokers have different origins, business models, and that there are multiple variations in how value is extracted from data. Data brokers are referred to in numerous ways, for example, as information brokers or resellers, consumer data collection companies, data aggregators, data providers, data suppliers or data intermediaries. Neither US nor European laws provide clear guidance on this. In a 2012 report, the US Federal Trade Commission (FTC) defined data brokers as “companies that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers for various purposes, including verifying an individual's identity, differentiating records, marketing products, and preventing financial fraud” (FTC, 2012, p. 68). In a 2014 report, the FTC provided a more concise definition: “companies that collect consumers’ personal information and resell or share that information with others” (FTC, 2014, p. i). In most cases, data brokers are considered companies that collect data through data mining techniques to create huge databases from which personal data can be extracted.

The data brokerage industry is very heterogeneous and complex. Companies like Acxiom, Oracle, Datalogix and Experian exchange information on most individuals in Western countries, yet these are likely unfamiliar names to most of their citizens. Nevertheless, these companies handle most of the internet data market, mainly composed of ad-tech groups, data analytics firms and credit agencies. Various customers, especially advertisers, employers, bankers, insurers, police departments and others, increasingly rely on the services provided by data brokers. Characterising the precise size of the data brokerage marketplace is difficult because of its vast scope and the variety of its operations. For example, both the US and EU still lack comprehensive lists or registries of such companies. Several privacy groups maintain lists of data brokers, but none are exhaustive or up to date (e.g. Privacy Rights Clearinghouse, 2020). Recent discussions are leading to public registries in the US (Abbott, 2019).

Several internet companies, advertisers, retailers, and trade associations sell personal data in various forms, and therefore could be considered data brokers (Kroft, 2014). In this sense, Facebook and Google may be among the biggest data brokers.[2] According to Grande (2014) Google and Facebook should be excluded from the above FTC definition of data brokers (FTC, 2014), which would otherwise be required to reveal more about their surreptitious information collection and use practices. They are, after all, first-party data miners: their data is primarily supplied by users who gave consent, rather than by other businesses (third-parties). Also, they could not have certain information that data brokers have. For example, many of Facebook's advertisers use behavioural profile data held by data brokers, rather than Facebook's own behavioural data, which is computed from actual user activity on the platform (Leetaru, 2018). Yet it is also relevant that big tech have a dominant role in the broader tracking ecosystem, owning several companies (West, 2019). For instance, Facebook and Datalogix have had significant partnerships (see Reitman, 2012; Shepherd, 2012). The relationship between big tech and data brokers remains complex and opaque. The inclusion of big tech in the definition of data brokers depends on how the latter is defined. Big tech indeed have substantial relations with data brokers, but they ultimately operate differently. Other companies, instead, often deny they are data brokers, though they have (self)identified as such to US regulators and data broker registries (Ng & Varner, 2021). More concise definitions and distinctions on data brokers' companies need to be elaborated upon (Sherman, 2021b).

2. In a sense, they may even be worse than data brokers who never presumed that you agreed to their practices (see Hoofnagle, 2003).

By tailoring their services for different purposes, various types of data brokers exist that sell products and services to various types of customers. The information, services and inferences they supply play central roles in key life decisions across a growing range of areas: a) advertising and marketing (e.g. micro-targeting or dynamic pricing), b) credit and insurance (e.g. for risk-mitigation[3]), c) identity verification and fraud detection (e.g. credit bureaus or people-search sites[4]), d) education, e) government and law enforcement and f) customer services. Then, information is sold at a relatively cheap price.

Data brokers collect very different types of data from various sources. Most often, they do not reveal the details of their data sources. The following is a non-exhaustive general list of sources:

- Provided by individuals. This occurs in several ways; most of the time when you use apps (e.g. games or weather apps) some data (GPS, audio, personal contacts etc.) is requested in order to access that service. Sometimes this can also occur with a "game-win strategy", in which people are persuaded to give out personal data in order to participate in a lottery (e.g. "subscribe and win a smartphone"). In some cases people share data with platforms, which in reality are secret data brokers.
- Cookies. Most online platforms use cookies to track and aggregate customers' overall activity across the whole internet.[5] The placement of such online cookies was pioneered by online advertisers, such as *DoubleClick*, which was acquired by Google in 2007. DoubleClick allows online advertisers to display different banners and even bid for the right to display their banners to a particular user (so-called real time bidding or programmatic advertising).[6] Users can usually opt out, even though most tend to accept cookies because of "information fatigue"[7] which results in

3. In the US, prospective employers already turn to data brokers to purchase criminal history reports regarding job candidates (reports that are notoriously error-prone). And police in both the United States and Europe purchase corporate assistance to profile residents based on personal data.
4. Websites such as *PeekYou* and *Spokeo* allow individuals and companies to find information about a person by searching for their name, phone number(s), address, email address and social-security number. This kind of services also allows what is referred to as "relational control", which occurs when individuals acquire the private data of those in their social or professional networks (see Rostow, 2017).
5. They can even use "web beacons" that by using a single-pixel GIF image, usually colored to match the background of a page or email - so that they are totally invisible - allow for the tracking of a tremendous amount of data on a user's behaviour: their typed entries and mouse movements, clickstream data, information from previously set cookies, and even recording conversations through a computer's microphone or images from the computer's camera (Sipior et al., 2011).
6. Apparently, in the course of auctions, the companies involved gain at least transient access to personal data, despite that the European Data Protection Regulation generally forbids companies from processing user data without consent. It is simply impossible for users to consent to real-time bidding when there is no way to know which companies are involved in an auction.

the "privacy paradox".[8]

- Software Development Kits (SDKs). Most often data is extracted by smartphone apps through SDKs. Usually, data brokers provide this software to developers for free. SDKs are used to make apps faster at the cost of allowing data brokers to hoard data (see Morrison, 2020).
- Third Parties. Some data brokers cooperate with third parties (e.g. ISP providers, online platforms, credit card networks operators, and other data brokers). Similarly, large online platforms, such as Facebook, Twitter and Google, cooperate with large data brokers to target ads and optimise their effectiveness.
- Government and public records. This is done by trawling public information generated by the State, such as property records, driver's licences, motor vehicle registrations, court records, census data, birth certificates, marriage licenses and voter-registration information, etc. Any data or information that is public is open to collection and exploitation, however privacy regimes might impede the use of this data in other contexts, such as marketing and propaganda.
- Data extraction through *web scraping* and *data crawling*. Basically, while web scraping refers to bots which crawl web pages simulating human Web surfing habits in order to collect specified bits of information from different websites, data crawling uses similar techniques to retrieve information from any source (not necessarily limited to the web). These techniques are legal, widely used, and almost impossible to avoid. Notably, along with public records, they can be used for "open-source intelligence".[9]

The information extracted from such data is manifold. Apart from demographic data (e.g. name, address, age, phone numbers, email, family ties, ethnic and religious affiliations, etc.), it is possible to extract other information such as "general interest data" (e.g. charitable giving, gambling, pets, preferred celebrities, movies and music genres, reading preferences, etc.), "home and neighbour data" (e.g. home equity, size, number of rooms and baths, rent price, loan amount and interest rate, etc.), "court and public data" (e.g. judgments, criminal offences, etc.), "social media and technology data" (e.g. internet provider, social media usage, operating system, mo-

7. Information fatigue refers to apathy, indifference, or mental exhaustion arising from exposure to too much information, especially stress induced by the attempt to assimilate excessive amounts of information from the media, the internet, or at work. Studies suggest that barely 30% of users regularly delete cookies (West, 2019).
8. The privacy paradox describes people's inconsistent willingness to protect their own privacy; people value privacy, but rarely act to protect it.
9. Open-source intelligence is 'produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement' (Williams & Blum, 2018). This could result in sensitive lists that could be used as a form of political intelligence, as was recently the case with a Chinese company who created a list of politically important networks in Western countries simply by using public information (see Balding, 2020).

bile devices usage and uploaded pictures, etc.), "financial data" (e.g. credit card usage, loans, net worth indicator, etc.), "health data" (e.g. tobacco usage, allergy sufferer, medicine preferences, etc.), "travel data", "vehicle data" and, last but not least, "purchase behaviour data" (FTC, 2014).

There is even other sensitive information that data brokers can stealthily deduct, such as problems with alcoholism (Hill, 2013), if you have just gone through a break up, depression, psychiatric problems and many others (see United States Senate Committee on Commerce, Science, and Transportation, 2013). All these pieces of information can be inferred through unsuspecting data, such as typing on a keyboard or swiping and tapping on a smartphone. The legal systems, however, are not always clear about (personal) data inferred from such datafied behaviours.

Section 2: The legal landscape and policy vacuum

The legal and political debates regarding data brokers – initially and mainly brokering credit scores – date back to the '70s.[10] For decades, data brokers operated in a market with little to no regulation, where transactions between corporations and governments were conducted without restrictions or public scrutiny (Crain, 2018). Almost anything was possible: capturing, buying, selling, sharing information, including that which was mined or statistically inferred. The development of the internet gave rise to an industry that provides an enormous number of services. Call for stronger regulation emphasised the privacy incursions of the data broker industry (Smith, 1997). Then, given various abuses in the past decade (most notably, the Cambridge Analytica scandal), as well as critical research on the risks big data have on human rights, public attention has turned again to data brokers, who are no longer as invisible as they once were. Indeed, data brokers practices, such as data-driven practices, can result in human rights infringements. They can inflict various types of harm on individuals, such as discriminatory profiling and have "chilling effects"[11] on expression and commerce. However, most studies have struggled to substantiate these concerns (Rieke et al., 2016) and therefore to push for tighter regulation.

The common argument used against data brokers is that consumers understand the bargain they make when conceding privacy for the benefits of using the inter-

10. See the Fair Credit Reporting Act of 1974, one of the first instances of data protection law passed in the computer age.

11. "Chilling effects" describe situations in which rights are threatened by possible negative results of exercising those same rights, namely it is a deterrent on the exercising of one's rights (see Büchi et al., 2020).

net (so-called privacy self-management[12]). In most cases, individuals can opt out from each data broker. This, however, requires a significant amount of time and gives limited certainty that data has been effectively deleted. These options are also often invisible and incomplete. For many commentators in the age of big data, providing people with a meaningful method to track all the data about them is practically impossible. As is well known, it would take a significant amount of time to read only the conditions of all the websites we visit – once calculated as 201 hours on average per year (McDonald & Cranor, 2008). It is indeed clear that privacy self-management has lost much of its effectiveness.

In the past decade, US institutions, including the Federal Trade Commission (FTC), the Government Accountability Office (GAO), and the Senate, have released reports and held hearings on the practices and operations of data brokers (FTC, 2014; Yeh, 2018). In 2012, the FTC proposed a tentative framework for privacy protection that businesses could adopt voluntarily and, when necessary, policymakers could employ it for general consumer protection. The framework included so-called "do not track" rules for web browsers to ensure user activity could be hidden from advertisers and allow for data portability capability, greater transparency, and consumer choice on where and how their data is shared with companies. Then, in 2014 the US Federal Trade Commission produced a report, compiled over two years, on nine of the biggest brokers. The Commission strongly recommended that Congress introduce legislation to limit the reach of data brokers.

Despite such discussions, data brokers remain mostly unregulated in the US. Many of them recognise virtually no rights for individuals in their policies, whereas some also include a clause that reserves the right of the company to change their data standards at any time. A number of legislative proposals have indeed been discussed. The *Information Transparency & Personal Data Control Act* would have required data brokers to get consent to collect sensitive data and pass through an annual privacy audit; the *Data Accountability and Trust Act* would have established security standards and require post-breach audits of data brokers and also prohibit collecting information under false pretences; the *Data Broker List Act 2019* would have required data brokers to sign up for a national registry overseen by the FTC and to maintain a comprehensive information security program as a means to protect consumer data from security breaches and other inadvertent or improper disclosures; finally, the *Data Broker Accountability and Transparency Act* of 2020 would have mandated opt-outs from data brokers and, again, for the FTC to create a na-

12. Privacy self-management is the idea that individuals can navigate, in a self-interested fashion, the complex balance involving privacy through rational decision-making and informed consent.

tional list of data brokers. None of these bills passed Congress. And while lobbying records do not always list specific bills, it is worth considering that in 2020 data brokers aggregate spending on lobbying rivalled the spending of individual Big Tech firms like Facebook and Google (Ng & Varner, 2021).

Unlike the United States, EU data protection legislation covers all private sector processing of personal data. EU legislation states that consumers have the right to access, correct, and object to the processing of their personal data. Furthermore, EU privacy legislation prohibits the processing of sensitive information unless an individual explicitly opts in to such processing, or the processing is allowed, as in specific cases listed in the GDPR.[13] This shields consumers from offensive or inaccurate data category classification, strengthening their position in the overall market.

EU legislation imposes ex ante control on the data controller. When collecting data, the controller must inform the consumer of the controller's identity and the reasons why the data are processed. In addition, it is forbidden to acquire more data than necessary (i.e. data minimization), which can also help protect consumer data from the risk of breach. The EU framework also imposes ex post control on enterprises, allowing consumers the ability to access, monitor, and correct personal data post-processing, as well as the ability to challenge data processing, such as the right to erasure(also referred to as the right to be forgotten, GDPR art. 17), the right to data portability, and the right not to be subject to a decision based solely on automated processing and profiling. Yet, certain controls are subject to commercial flexibility exceptions, which may undermine privacy protection. Finally, EU legislation provides strong sanctions and compensation that incentivise companies to take regulation seriously.[14]

On both sides of the Atlantic, there is not yet a clear regulatory agenda for data brokers. Not only because of lobbying pressure, but also because all parties involved – first parties (e.g. Google), second parties (e.g. Samsung), and third parties (e.g. data brokers) – have a mutual stake in circumventing policy by building new data extraction techniques. And although there is ample reason to be concerned about these data flows, it is also assumed that many of these activities pose minimal risk to human rights. Many commentators have argued that it is sufficient to

13. GDPR recital 51.

14. Those who infringe upon certain provisions of the GDPR can face administrative fines of up to 20 million EUR, or up to 4% of the total worldwide annual turnover from the preceding financial year, whichever is higher (GDPR art. 83(4) & (5)). Consumers also have the right to receive compensation from the controller or processor for the damage suffered (GDPR art. 82).

regulate data brokers under a comprehensive legal framework, taking the European one as a model (e.g. Kuempel, 2016; Yeh, 2018). While the GDPR model provides new rights to consumers that might ultimately temper the risks arising from data brokers activities, it does not address the underlying role data brokers play in surveillance capitalism. Robust-sounding legal principles are established, but public authorities and civil society often struggle to apply them in concrete ways (Rieke et al., 2016). Furthermore, the effectiveness of the GDPR's above provisions is questionable, as they remain a more recent regulation yet to be fully assessed in practice. Compliance with basic provisions of the GDPR in apps on the Google Play Store, for example, is limited and, despite transparency rules, the analysis of privacy practices in the mobile tracking ecosystem remains difficult (Kollnig et al., 2021). Also, the GDPR tends to focus on browser-based (e.g. cookies), which are not app-driven technologies (e.g. SDKs) and allow most data brokers to collect data (Morrison, 2020). These concerns may also hold true for other US regulations, such as California's CCPA Data Act, as well as state laws (such as Virginia's and Maine's privacy laws). Hopefully, forthcoming regulations, like the European Data Governance Act, could develop stricter policy solutions. Current approaches, however, do not predict more restrictive rules and strong oversight. There are indeed important steps forward, such as the creation of registries or annual privacy impact assessments (only for large data brokers), yet the approach remains rather soft, mostly oriented to individual empowerment and with a limited global focus. The next sections thus develop a more critical approach to data brokers, stressing their political role and introducing a number of systemic policy proposals to develop more comprehensive legislation.

Section 3: The role of data brokers in surveillance capitalism

More than just information resellers, data brokers enable information exchange among organisations and, eventually, create markets for consumer data which, as Crain (2018) argues, further incentivises surveillance among many types of entities. By doing so, they also reproduce and extend the processes of audience commodification, which is deeply entrenched in historical processes of capitalist expansion (Andrejevic, 2010). Users are considered commodities. Their activity on the web is unconscious work for the benefit of internet companies.[15] Data brokers

15. The commodification of users is ultimately calculated with the metric *Average Revenue Per User* (so-called ARPU). It is defined as the total revenue divided by the number of subscribers/users of social media platforms, Internet Service Providers and other companies. It is indeed a rough estimation as to what one's personal data might be worth. Yet, the future reuse value of the data made by data brokers is usually not accounted for. The very secrecy of this industry impedes a reliable estimation.

have thus constructed an environment in which individuals are "constantly surveyed and evaluated, investigated and examined, categorised and grouped, rated and ranked, numbered and quantified, included or excluded, and, as a result, treated differently" (Christl, 2017). Therefore, data brokers are fundamental actors of surveillance capitalism since they engage in a sort of "information arbitrage": buying, reinterpreting, repackaging, and selling consumer data across contexts. Such an organisational, exploitative and pervasive role cannot be understated. There are also serious concerns that the pandemic and the technologies employed to combat it, such as digital identity systems, vaccine passports, and border crossing apps, have actually broadened opportunities for data brokers to collect data (Mills Rodrigo, 2021).[16]

In some cases it is easier to purchase detailed data about a population from data brokers than it is to request the same kind of data from the government. In addition, purchased data cannot be further examined or corroborated because of the data brokers' intellectual property protections. This can increase data inequality. For example, well-funded researchers or entities in collaboration with data brokers will have more opportunities to publish research than less well-funded researchers or the general public. From an economic perspective, Muralidhar and Palk (2018) have explored how data brokers further inequality in accessing credible data through a rent-seeking behaviour,[17] which is considered detrimental to a free-market economy. Data brokers will likely contend that they are not rent seekers, not as long as they provide added value by aggregating datasets through independently created algorithms, thus allowing third parties to develop a fuller picture about consumers and providing them with relevant information. In this regard, data brokers simply access free data from individuals, who are generally unaware that their data is being repurposed and sold to third parties to convince users to purchase products they might not have otherwise bought. This process is much more effective than in traditional advertising. It arguably represents a transfer of wealth rather than the creation of wealth, and it could be exacerbated if and when "data altruism" initiatives, such as the "European Data Spaces", are implemented.

16. For a website which analyses the technologies developed for the pandemic see Tactical Tech (2020).

17. Rent-seeking is a theory of economic behaviour that entails asking the government for certain privileges or deriving significant profits and advantages without adding any value to the economy. More simply, it consists of transferring wealth rather than creating wealth. This behaviour is criticised as contributing to economic inefficiency and economic inequality, as the wealthy receive the benefits of anticompetitive rent-seeking behaviour while the rest of the market suffers the losses.

Effectively opting out from data brokers' data collection is time-consuming and does not leave certainty that this optout includes all personal data; this, in turn, leads to question whether users can effectively and autonomously exercise their right of erasure. Other than free websites, which support users by giving them the chance to opt out,[18] there are also optout services such as Kanary and Deleteme who search through data brokers, as well as a list of social media platforms, people search sites, and search results for pages exposing personal information to the public. Yet, they also admit that they cannot guarantee information removal. The truth is that once personal information has been packaged, sold and resold, it may live indefinitely in the servers run by the data broker industry. If it is hacked – and it more likely has been[19] – then the profile joins the billions of other profiles being traded on the dark web. This represents another concerning feature of the data broker industry: virtually endless profiling persistence. The right of erasure is in fact essential, but its effectiveness is debatable; among several reasons, it is unclear under what conditions data processing can be considered "unlawful" (under the Article 17 of the GDPR).

Similarly, it is extremely difficult to escape data brokers' surveillance. Data capturing is ever more pervasive, especially on social media; by auditing data brokers via Facebook's advertising platform, for example, Venkatadri et al. (2019) found that a surprisingly large percentage of Facebook accounts (e.g. above 90% in the US) are successfully linked to data broker information. Moreover, there are increasingly sophisticated techniques of re-identification. Data brokers and similar companies provide cross-device tracking services that are based on using machine learning to analyse large amounts of data.[20] There are also techniques and tools (e.g. AdNauseam or VPNs) that help you to obfuscate your digital footprints. Yet, it is questionable whether the average user is aware of these tools and willing to use them.

In the data broker industry, but also more generally in the big data industry, there is a widespread assumption that data is self-explanatory, and that big data alone always results in more predictive power. However, any data correlation is useful only under certain background assumptions, which ultimately come from theory.

18. For example, the website *yourdigitalrights.org* offers a guide to sending GDPR Erasure Requests.
19. See, for example, if your email and password have been leaked on *haveibeenpwned.com*. You can also check data breach search engines such as *dehashed.com*.
20. For example, the company *Tapad* analyses data on 2 billion devices around the globe and uses behavioural and relationship-based patterns to find the statistical chance that certain computers, tablets, phones and other devices belong to the same person. Similarly, data broker Acxiom offers LiveRamp IdentityLink, an identity graph that matches directly identifiable data – like emails, postal addresses, and phone numbers – with pseudonymous identifiers – like cookies and device IDs.

Without it, correlations can be as misleading as they are informative (i.e. apophenia). This positivistic process of constructing meaning is exceptionally political (Kitchin, 2014) and it can easily lead to data inaccuracy and, accordingly, inferences that can be wrong and potentially discriminatory. Data brokers can, for example, adjust prices based on anticipated behaviour (i.e. dynamic pricing), typecast someone as a consumer with a bad credit record, or as a person with health problems that could affect job performance, even if the information on which these notions are based is incorrect. Venkatadri et al. (2019) showed that at least 40% of data broker sourced user attributes on Facebook are not at all accurate, even when it came to financial information. Data brokers have even reportedly sold lists of rape victims, alcoholics and erectile dysfunction sufferers. Or, they have segmented consumers with biased labels, for example, "rural everlasting" (single men and women over the age of 66 with little education and small net worth) (Anthes, 2015). Thus, it could be fundamental to afford individuals and the broader society the chance to contest these inferences. Nevertheless, this may not even be enough if the incentives that make such inferences valuable in the first place are not disrupted.

As a matter of fact, data brokers will never achieve meaningful transparency – for example by releasing comprehensive information about their practices and access to individual databases – because, as Crain (2018) explains, the structure and operations of the industry are naturally incompatible with a transparency framework of full disclosure. While the intentions of transparency are fair, it is a policy approach that is subsumed by a discourse of consumer empowerment that has been rendered meaningless in the contemporary environment of pervasive commercial surveillance: privacy asymmetry is indeed a cornerstone of the data broker business model. Moreover, data brokers' appropriate transparency values public-relations efforts to deflect the threat of government regulation. Transparency initiatives have historically been deployed to advocate regimes of industry self-regulation, which have repeatedly failed to protect consumer privacy (Crain, 2018). This ultimately follows a traditional pattern in internet history whereby commercial surveillance is legitimised by the illusion of consumer choice.

Most of the data brokers' market is concerned with marketing and online advertisement (Christl, 2017). Supporters of data brokers' role in society claim, such as, unsurprisingly, the chief data ethics officer for data broker company Acxiom, Jordan Abbott (2019), that marketing data brings real value to consumers; advertising helps consumers receive relevant information and assists them in making choices on an endless array of goods and services. Similarly, it helps sellers to understand

customers and deliver them marketing messages that are more relevant, consistent and effective. And in many cases data marketing even "funds the press and other channels of expression" (ibid) while, instead, for many commentators it is the ad-based revenue model to incentivise sensationalistic journalism, clickbait and, overall, to negatively affect the quality of the press. More often, misinformation and conspiracy theories are the product of this business model, not an accident (Kingaby & Kaltheuner, 2020). The fundamental role of data brokers in online advertising deserves further investigation.

Despite a lack of evidence, data brokers could also have an important role in political campaigns and cyber espionage (Leong & Yi-Ling, 2020), representing a weak link in national security (Twetman & Bergmanis-Korats, 2021; Sherman, 2021a). In the hands of malicious actors, data can indeed become a tool for disinformation operations, for example during elections (Anstead, 2017). Think again of the Facebook/Cambridge Analytica scandal, or more recently, of the TikTok ban in the US due to concerns that it could be used for surveillance or espionage by China; cyber policy scholar Samm Sacks argues that American companies can still sell data to data brokers, even after buying ownership of foreign-based apps (Roose, 2020). Though at a high price, data brokers could then turn around and sell data to companies that, in turn, may sell it to the Chinese government. They can indeed proactively obfuscate the source of their data, making it difficult for anyone to retrace the paths through which the data was collected. Data brokers are indeed incentivised to develop software-driven strategies to circumvent any privacy law. They can do this with the protection of trade secrets, non-disclosure and even non-disparagement agreements, for example, to stop former employees from whistleblowing. All the above considerations indicate the nontrivial political role of data brokers and raise serious concerns on the effectiveness – or even enforceability – of data protection regulation at the global level.

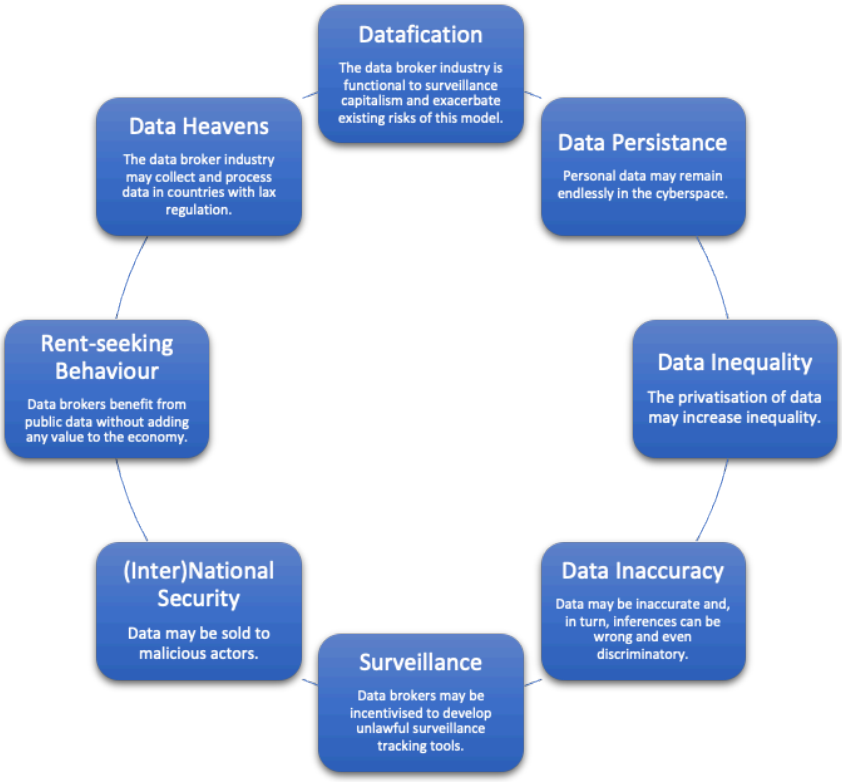

FIGURE 1: Main risks of an under-regulated data brokers industry.

Section 4: Policy Perspectives

The complexity and variety of the data brokerage industry do not allow for simplifications and generalisations. There are, nonetheless, a number of debates that need to be elaborated further to understand how to tackle the challenges described above. Generally speaking, forthcoming regulations mainly focus on palliative solutions that could never be fully effective due to the fact that data brokers' purposes are deeply entrenched in the current paradigm of surveillance capitalism. This should not be underestimated. There are fundamental systemic challenges at stake in the emerging global data economy, and these need to be developed in order to effectively regulate the data broker industry. "Data governance" is still a fuzzy notion, and no single regulation governs the subject in a comprehensive fashion. Multiple regulatory fields have to be carefully developed, especially to avoid unintended effects among different domains.

The fluidity of data and its protection across borders: standardisations and inferences

Data is multidimensional, fluid and has unusual properties. Its features challenge its classification and, as a consequence, its regulation and protection as well. Data

is, above all, a non-rivalrous resource that can be replicated and combined in numerous value chains without being depleted, yet it is an excludable one. As such, data is often fragmented or copied, with different component parts stored in different places, they are constantly being disassembled and reassembled, and moved across servers. Thus, users usually have no control over where their data is stored and oftentimes lack any knowledge of where it is located, and which jurisdiction might govern it. Fundamentally, data challenges the role of territory as the basic defining spatial epistemology in international law. Moreover, corporations have powerful roles in this domain, as they are the ones who decide where to store data, where to establish headquarters, in which jurisdictions to establish data centres, on what person(s) they collect data, and how to mediate disputes over data across borders. As such, the legal status and theoretical conceptualisations of data and its governance still need to be properly developed. Notably, there are cases in which it is still unclear if information inferred by algorithms about a person falls within the categorisation of personal or sensitive data. This is an accountability gap even in the GDPR. In theory, if all data has a potential to impact people, then all data could be considered personal or sensitive and thus would need protection (Purtova, 2018). The world of data is indeed prone to overlapping modes of classification and formed by numerous actors connected by complex relations and value chains. The multi-dimensional nature of classifications is a serious obstacle to universal taxonomies and simple-to-implement rules. It is therefore essential to standardise global and international laws aimed at protecting individuals while favouring the competition and downsizing the monopolist positions held by few players in specific markets. Finally, in addition to data standardisation and protection against input data, it is essential that individuals are also protected against outputs of data processing, namely the potential harms that could result based on inferences. A "right to reasonable inferences" could therefore potentially address the accountability gap of data brokers that, eventually, would have an obligation, ex ante, to justify the reasonableness of an inference (Wachter & Mittelstadt, 2019).

Privacy as a relational common good: group privacy

Due to a set of externalities and information leakages involved in data markets, privacy can actually be recognised as an "aggregate public good" prone to market failure. Recognizing this fact should convince us that government intervention is both beneficial and necessary for its protection (Sætra, 2020). This is even more relevant considering the flaws of privacy self-management and emerging algorithmic techniques that might threaten the privacy of groups. In fact, algorithmic systems which measure, count, and profile groups of individuals create knowledge

that is not (only) private to an individual, but which reveals details about a group of individuals. This is framed as the protection of "group privacy" (Taylor et al., 2016). Privacy researchers have long proposed a contextual and relational understanding of privacy, mainly referred to as relational or contextual privacy (Nissenbaum, 2011; Bannerman, 2019). As Viljoen (2020) argues, "the data collection practices of the most powerful technology companies are primarily aimed at deriving population-level insights from data subjects that can then be applied to individuals that share these population features, not individual-level insights specific to the data subject in question" (p. 3). Such awareness ultimately calls for overcoming the individual privacy self-management paradigm that has legitimised commercial surveillance so as to move towards a "relational data governance". To begin, the informed consent paradigm should be reformed. For example, the content of contractual default provisions could depend on the articulated preferences of ordinary consumers as measured by scientifically rigorous survey instruments (i.e. "consumertarian" default rules[21]) (Stigler Center for the Study of the Economy and the State, 2019). As a result, terms and conditions would by default preserve more privacy. More generally, experts have already developed several ways to protect privacy more effectively, ranging from privacy-preserving systems such as "differential privacy"[22] to privacy impact assessments.

The challenges of personal data governance

New models to personal data governance have been developed and implemented (Ritter & Mayer, 2017; Mills Rodrigo, 2021). In general, three major models of personal data governance are usually discussed: laissez faire, data commons and data trusts. Laissez faire is the current mainstream model that has been analysed and criticised in this article. Alternative models such as data trusts (Delacroix & Lawrence, 2019), data cooperatives (Hardjono & Pentland, 2019), personal data stores, data commons and semi-commons, tend to increase individual control and, therefore, arguably better tackle risks arising from the data broker industry. Cofone (2021), instead, considers these models inadequate in protecting privacy rights and proposes to reinforce, above all, the "purpose limitation principle". While privacy harm can be produced at the moment of collection, processing, or dissemination of

21. Consumertarian Default Rules are default rules on data protection that follow the preferences of a majority of consumers (Stigler Center for the Study of the Economy and the State, 2019). These are supposed to be revisited periodically to account for updates in consumers' preferences.

22. Differential privacy is a formal mathematical framework for quantifying and managing privacy risks against a wide range of potential attacks. It applies to analyses of collections of individual information and, therefore, is particularly suitable to protecting "group privacy"; it helps to overcome the limitations of earlier anonymization techniques (Feldman et al., 2020).

personal information, property rules can only control the moment of collection, he argues, and eventually produce a moral hazard problem. Unless otherwise constrained, companies lack incentives to minimise processing and disclosure harms after data has been exchanged. In any case, these personal data governance frameworks are competing with each other for adoption by citizens, public and private stakeholders, and it is fundamental to understand how macro-frameworks (i.e. GDPR) can shape and favour meso-level governance logic (i.e. personal data governance frameworks) (Bodó et al., 2021). The issue of data ownership in the light of the data brokerage industry is contentious and undoubtedly deserves more discussion.

A new social contract for data and cybersecurity?

In light of the above, it is fundamental to explore and eventually combine alternative forms to govern not just data, but also the private infrastructures that allow data brokers to collect, process and exchange data. Developing a set of internationally agreed upon principles for the regulation of data brokers seems a complex task for broad-based negotiations, such as the G20 and WTO. Convergence might instead be more likely within smaller groups of like-minded countries.[23] Yet, organising a global multistakeholder debate across sectors could improve both awareness and agreement (De La Chapelle & Porciuncula, 2021). Similar to the Internet Governance Forum, this has been advocated to ensure that not only all states, but also other stakeholders, such as the private sector, civil society and technical community, can equally participate in designing, developing and ultimately implementing any proposed approach. In turn, this endeavour would help to add nuance and rebalance a debate that is currently polarised, generally, between those in favour of the "free flow of data" and those in favour of "data sovereignty". This might also help to facilitate a creative discussion that is more global, evidence-based, and focused on common objectives, as well as to explore innovative approaches in tools, frameworks and concepts for dealing with data. This dialogue could lead to a new "social contract for data".[24] Such prospective agreement would also have to tackle cyber-security governance issues. In the last few years,

23. For example, at the end of October 2021 Trade Ministers of the G7 countries issued a set of commonly agreed upon Digital Trade Principles (see UK Department for International Trade, 2021).

24. Proposals that move in this direction are, for example, the concept of "data altruism" developed in the forthcoming European Data Governance Act – in which individuals or companies make data voluntarily available for reuse, without compensation, for the common good – or the creation suggested by the World Bank (2021) of an integrated national data system (INDS) – that would integrate participants from civil society and the public and private sectors into the data life cycle and into the governance structures of the system, allowing the flow of data among a wide array of users in a way that facilitates safe use and reuse of data.

in fact, data breaches have steadily increased and several countries are even suspected of "government hacking", which means they used sophisticated data mining technology, including spyware. Not only data brokers would benefit from these tools, but this also reveals the need for a discussion on the regulation of the broader private surveillance market.

A new oversight authority?

A prospective outcome of this could be the establishment of a new, independent, transnational, authority. In this context, an academic and policy committee organised by the Stiegler Center (2019) proposed the creation of a digital authority: a single powerful regulator capable of overseeing all aspects of digital platforms and, similarly, data brokers. The purpose is to generate several concerns across different fields, all linked to the power of data. To address these concerns in a holistic way, there needs to be a single regulator able to, among many others, impose open standards, to mandate portability of and accessibility to data, and eventually, to monitor data brokers and other actors in surveillance and data markets. This would certainly require a careful institutional design to preserve transparency and to avoid being captured by the industry. Along a similar line, new legal devices could be employed. For example, data centres could be legally claimed as "critical information infrastructures",[25] subject to more stringent security regulation, such as effective cybersecurity measures to protect data and uphold transparency and accountability (Leong & Yi-Ling, 2020). To reach these bolder policies, it is preconditional, however, to effectively counteract the evident lobbying that has successfully allowed the policy vacuum described above.

Fix the ad-industry?

Eventually, the advertisement business model that currently sustains most internet services and is the core business of data brokers, could be reformed. For many reasons – among them ad-blockers, out-of-sights ads and click-farms – the effectiveness (and thus returns) of online ads can be seriously questioned (Neumann et al., 2019). Not only is there little evidence that constant tracking leads to more relevant ads, but a recent study showed how the availability of cookies increases publisher's revenue by only about 4% (Marotta et al., 2019). Unsurprisingly, Google has argued that publishers would lose half their revenue or more if they stopped using

25. Critical Information Infrastructure are those interconnected information and communication infrastructures which are essential to the maintenance of vital societal functions, (health, safety, security and the economic or social well-being of people) – the disruption or destruction of which would have serious consequences.

personalised advertising (Ravichandran & Korula, 2019). Others argue that online ads are so overvalued that they might even represent the next financial bubble (Hwang, 2020). A compelling policy approach is proposed by the Nobel Prize winner for economics Paul Romer (2021), who argues to enact a progressive, sufficiently aggressive tax on revenue from digital advertising. This could make the subscription model more attractive or, more simply, make it more attractive for a large firm to create independent new ventures, and less attractive for it to grow via acquisitions. Another debated policy proposal is the ban of 'surveillance ads', not only for their undesirable consequences, but because they appear to have limited effectiveness (Edelman, 2020). This ban can take different forms. One of the most advocated is a ban on the use of personal data (in particular psychometric data) for "behavioural advertising", and to opt instead for "contextual advertising", which basically depends on the content of the web page the user is viewing (Gary & Soltani, 2019). Such a ban could help to protect individual privacy, reduce corporate incentives to maximise invasive data collection and spur innovation in the advertising sector.

Research directions

The relevance and complexity that the data brokerage industry entails deserves further investigation, not only from journalists and academics but, in particular, from regulators and policy-makers. Generally speaking, a sustainable environment for data brokers is one in which regulation is enforced globally, at least at European standards, and the negative externalities of data brokers minimised while privacy by design is enforced from data infrastructures to users' daily habits (e.g. informed consent and privacy awareness). Importantly, the management and legal status of data needs further conceptual elaboration and standardisation. Similarly, new definitions that narrow down similar, but different data brokers' subjects, are critical (see Figure 2). For example, what are the differences between standard data brokers, quasi-data brokers like Big Tech, and unregistered data broker markets on the deep web? Their activities? Their business models? And, then, what are the relationships between data brokers, Big Tech, and the ad-industry? How might we go beyond the current practice of informed consent to protect privacy collectively? How can we fully guarantee data centres' cybersecurity, data developers transparency and fair profiling processes in this context? How can we quickly detect human rights breaches, if any exist? And what tools and efforts exist to make data brokerage-driven data harvesting readable and empirically criticisable? To answer these questions convincingly, it is undoubtedly paramount that social scientists collaborate with engineers and programmers.

Other more speculative questions can also be raised in this fast-changing context; To what extent are data brokers technically able to re-identify users online? Do data brokers really have the ability to effectively obfuscate their data sources and stealthily move data across the globe? If these concerns were real, it is even possible to expect incentivised subjects to create data broker companies that are even less accountable, more decentralised, and ultimately able to bypass oversight and regulations? What is (and what could be), then, the role of data brokers in the (geo)political global arena? Eventually, could data brokers proactively create, similar to fiscal havens, 'data havens' where privacy regulations are much less stringent? In that case, what would be the consequences? These concerns are concrete and deserve not only further research, but a series of investigations that no specific global authority seems responsible for, or is currently qualified to launch.

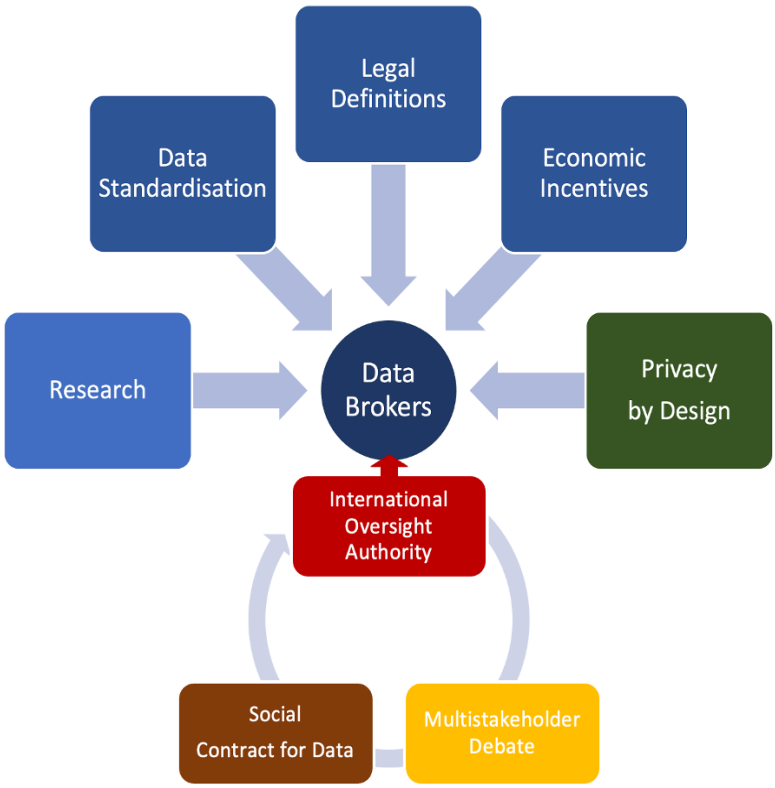

FIGURE 2: Summary of the main policy approaches for the regulation of data brokers.

Conclusion

Despite being politically underestimated and legally under-regulated, the data broker industry has a fundamental role in surveillance capitalism. Data brokers collect enormous amounts of personal data, most often in stealthy ways. Then, they analyse such data – while assessing, rating and judging individuals – so as to extract information to sell to other companies, private individuals and even (for-

eign) governments. Data is also often exchanged with third parties (including Big Tech) and effective consumer protection rules are difficult, not only to enact, but also to enforce. Eventually, among the thousands of data brokers worldwide, there are serious risks of data breaches whose (often inaccurate) inferences could lead to discrimination, manipulation and even to (inter)national security issues.

The data broker industry legitimises society's datafication, allowing the potentially endless persistence of personal data in cyberspace. By its nature, it cannot be fully transparent and, at the same time, it has the ability to circumvent accountability measures. These considerations may even lead to questions about the extent to which such an industry can ever be compatible with privacy (and democracy) at all. There is little doubt that the data broker industry deserves more scrutiny, and that privacy should be protected first and foremost by default, not so much individually. It is likely, in fact, that the role of data brokers will continue to become more and more invasive if left unchecked, as it currently is. So far, regulation seems timid and unable to tackle privacy threats at a global and collective level. Still, there is no seriously innovative and ambitious policy agenda. The forthcoming Data Governance Act in Europe, for example, could represent an important opportunity to implement more comprehensive rules. The challenges of regulating data brokers are in fact deeply entrenched in the challenges of data governance, international law and surveillance capitalism. Far from a full understanding of the subject – mainly due to its opaque, transnational nature, as well as its heterogeneous activities – the author's goal is to stimulate further research and discussion, taking into account the systemic, economic and political role of the data broker industry so as to ultimately scrutinise it with a more proactive, multidisciplinary and transnational approach, rather than one that is reactive, sectorial and adaptive.

References

Abbott, J. (2019, September 13). Time to build a national data broker registry. A federal clearing house would help separate the good and bad actors who deal in data. *The New York Times*. https://www.nytimes.com/2019/09/13/opinion/data-broker-registry-privacy.html

Andrejevic, M. B. (2010). Surveillance and alienation in the online economy. *Surveillance & Society*, *8*(3), 278–287. https://doi.org/10.24908/ss.v8i3.4164

Anstead, N. (2017). Data-driven campaigning in the 2015 United Kingdom general election. *The International Journal of Press/Politics*, *22*(3), 294–313. https://doi.org/10.1177/1940161217706163

Anthes, G. (2015). Data brokers are watching you. *Communications of the ACM*, *58*(1), 28–30. https://doi.org/10.1145/2686740

Balding, C. (2020). Chinese open source data collection, big data, and private enterprise work for state intelligence and security: The case of Shenzhen Zhenhua. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3691999

Bannerman, S. (2019). Relational privacy and the networked governance of the self. *Information, Communication & Society*, *22*(14), 2187–2202. https://doi.org/10.1080/1369118X.2018.1478982

Bodó, B., Irion, K., Janssen, H., & Giannopoulou, A. (2021). Personal data ordering in context: The interaction of meso-level data governance regimes with macro frameworks. *Internet Policy Review*, *10*(3). https://doi.org/10.14763/2021.3.1581

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., Velidi, S., & Viljoen, S. (2020). The chilling effects of algorithmic profiling: Mapping the issues. *Computer Law & Security Review*, *36*, 105367. https://doi.org/10.1016/j.clsr.2019.105367

Christl, W. (2017). *Corporate surveillance in everyday life: How companies collect, combine, analyze, trade, and use personal data on billions* [Report]. Cracked Labs - Institute for Critical Digital Culture. https://crackedlabs.org/en/corporate-surveillance

Cofone, I. (2021). Beyond data ownership. *Cardozo Law Review*, *43*(2), 501–572. https://doi.org/10.2139/ssrn.3564480

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88–104. https://doi.org/10.1177/1461444816657096

De La Chapelle, B., & Porciuncula, L. (2021). *We need to talk about data: Framing the debate around free flow of data and data sovereignty* [Report]. Internet and Jurisdiction Policy Network. https://www.internetjurisdiction.net/uploads/pdfs/We-Need-to-Talk-About-Data-Framing-the-Debate-Around-the-Free-Flow-of-Data-and-Data-Sovereignty-Report-2021.pdf

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Edelman, G. (2020, March 22). Why don't we just ban targeted advertising? *WIRED*. https://www.wired.com/story/why-dont-we-just-ban-targeted-advertising

Federal Trade Commission. (2012). *Protecting consumer privacy in an era of rapid change* (FTC Report March 2012). https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

Federal Trade Commission. (2014). *Data broker: A call for transparency and accountability* (Federal Trade Comission May 2014). https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Feldman, V., Kakaes, K., K., L., Nissim, K., Slavkovic, A., & A, S. (2020). *Differential privacy: Issues for policymakers* [White paper]. Simons Institute for the Theory of Computing. https://simons.berkeley.edu/news/differential-privacy-issues-policymakers

Gary, J., & Soltani, A. (2019). *First things first: Online advertising practices and their effects on platform speech* (Free Speech Futures) [Essay]. Knight First Amendment Institute at Columbia University. https://knightcolumbia.org/content/first-things-first-online-advertising-practices-and-their-effects-on-platform-speech

Grande, A. (2014, March 14). FTC's brill excludes Google, Facebook from data broker push. *Law360*.

http://www.law360.com/articles/518639/ftc-s-brill-excludes-google-facebook-from-databroker-push

Hardjono, T., & Pentland, A. (2019). *Data cooperatives: Towards a foundation for decentralized personal data management* (1905.08819). arXiv. https://arxiv.org/pdf/1905.08819.pdf

Hill, K. (2013, September 19). Data broker was selling lists of rape victims, alcoholics, and 'erectile dysfunction sufferers'. *Forbes*. https://www.forbes.com/sites/kashmirhill/2013/12/19/data-broker-was-selling-lists-of-rape-alcoholism-and-erectile-dysfunction-sufferers

Hoofnagle, C. J. (2003). Big brother's little helpers: How ChoicePoint and other commercial data brokers collect and package your data for law enforcement. *N.C. J. Int'l L.*, *29*(4), 595–638.

Hwang, T. (2020). *Subprime attention crisis: Advertising and the time bomb at the heart of the internet*. Farrar, Straus & Giroux.

Kingaby, H., & Kaltheuner, F. (2020). *Ad break for Europe: The race to regulate digital advertising and fix online spaces* [Policy brief]. Mozilla Foundation. https://assets.mofoprod.net/network/documents/Ad_Break_for_Europe_FINAL_online.pdf

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. https://doi.org/10.1177/2053951714528481

Kollnig, K., Binns, R., Van Kleek, M., Lyngs, U., Zhao, J., Tinsman, C., & Shadbolt, N. (2021). Before and after GDPR: Tracking in mobile apps. *Internet Policy Review*, *10*(4). https://doi.org/10.14763/2021.4.1611

Kroft, S. (2014, August 24). The data brokers: Selling your personal information. *CBS News*. https://www.cbsnews.com/news/data-brokers-selling-personal-information-60-minutes

Kuempel, A. (2016). The invisible middlemen: A critique and call for reform of the data broker industry. *Nw. J. Int'l L. & Bus*, *36*(1), 207–234.

Lazarus, D. (2019, November 5). Column: Shadowy data brokers make the most of their invisibility cloak. *Los Angeles Times*. https://www.latimes.com/business/story/2019-11-05/column-data-brokers

Leetaru, K. (2018, April 5). The data brokers so powerful even Facebook bought their data — But they got me wildly wrong. *Forbes*. https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong

Leong, D., & Yi-Ling, T. (2020, August 21). Data brokers: A weak link in national security. *The Diplomat*. https://thediplomat.com/2020/08/data-brokers-a-weak-link-in-national-security

Marotta, V., Abhishek, V., & Acquisti, A. (2019). Online tracking and publishers' revenues: An empirical analysis. *WEIS Workshop on the Economics of Information Security*. https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_38.pdf

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, *4*(3), 540–565.

Mills Rodrigo, C. (2021, November 10). Data broker shared location data with DC government for coronavirus tracking. *The Hill*. https://thehill.com/policy/technology/580975-data-broker-shared-location-data-with-dc-government-for-coronavirus

Morrison, S. (2020, July 8). The hidden trackers in your phone, explained: How covert code enables your phone's apps to spy on you. *Vox*. https://www.vox.com/recode/2020/7/8/21311533/sdks-tracki

ng-data-location

Muralidhar, K., & Palk, L. (2018). A free ride: Data brokers' rent-seeking behavior and the future of data inequality. *Vanderbilt Journal of Entertainment and Technology Law*, *20*(3), 779–837.

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? Evidence from field studies. *Marketing Science*, *38*(6), 918–926.

Ng, A., & Varner, M. (2021, April 1). The little-known data broker industry is spending big bucks lobbying congress. *The Markup*. https://themarkup.org/privacy/2021/04/01/the-little-known-data-broker-industry-is-spending-big-bucks-lobbying-congress

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, *140*(4), 32–48. https://doi.org/10.1162/DAED_a_00113

Privacy Rights Clearinghouse. (2020). *Data Breaches* [Database]. https://privacyrights.org/data-breaches

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, *10*(1), 40–81. https://doi.org/10.1080/17579961.2018.1452176

Ravichandran, D., & Korula, N. (2019). *Effect of disabling third-party cookies on publisher revenue*. Alphabet Inc. https://services.google.com/fh/files/misc/disabling_third-party_cookies_publisher_revenue.pdf

Reitman, R. (2012, September 25). *A deep dive into Facebook and Datalogix: What's actually getting shared and how you can opt out*. Electronic Frontier Foundation. https://www.eff.org/it/deeplinks/2012/09/deep-dive-facebook-and-datalogix-whats-actually-getting-shared-and-how-you-can-opt

Rieke, A., Yu, H., Robinson, D., & Hoboken, J. von. (2016). *Data brokers in an open society* [Upturn Report]. Open Society Foundations. https://www.opensocietyfoundations.org/publications/data-brokers-open-society

Ritter, J., & Mayer, A. (2017). Regulating data as property: A new construct for moving forward. *Duke Law & Technology Review*, *16*(1), 220–277.

Romer, P. (2021, May 17). *Taxing digital advertising*. Adtax. https://adtax.paulromer.net

Roose, K. (2020, July 26). Don't ban TikTok. Make an example of it. *The New York Times*. https://www.nytimes.com/2020/07/26/technology/tiktok-china-ban-model.html

Rostow, T. (2017). What happens when an acquaintance buys your data: A new privacy harm in the age of data brokers. *Yale Journal on Regulation*, *34*(2), 667–707.

Sætra, H. S. (2020). Privacy as an aggregate public good. *Technology in Society*, *63*, 101422. https://doi.org/10.1016/j.techsoc.2020.101422

Shepherd, T. (2012, November 12). Desperation and Datalogix: Facebook six months after its IPO. *Culture Digitally*. https://culturedigitally.org/2012/11/desperation-and-datalogix/

Sherman, J. (2021a). *Data brokers and sensitive data on U.S. individuals. Threats to American civil rights, national security, and democracy* [Report]. Duke University. https://sites.sanford.duke.edu/techpolicy/report-data-brokers-and-sensitive-data-on-u-s-individuals/

Sherman, J. (2021b, April 8). Federal privacy rules must get "data broker" definitions right. *Lawfare*. https://www.lawfareblog.com/federal-privacy-rules-must-get-data-broker-definitions-right

Sipior, J. C., Ward, B. T., & Mendoza, R. A. (2011). Online privacy concerns associated with cookies, flash cookies, and web beacons. *Journal of Internet Commerce*, *10*(1), 1–16. https://doi.org/10.1080/15332861.2011.558454

Smith, R. E. (1997, February 1). Privacy: The untold stories. *WIRED*. http://www.wired.com/1997/02/cyber-rights-13

Stigler Center for the Study of the Economy and the State. (2019). *Stigler Committee on Digital Platforms* [Final report]. https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf

Tactical Tech. (2020). *Technologies of hope & fear: 100 responses to the pandemic*. https://tacticaltech.org/news/techpandemic

Taylor, L., Floridi, L., & van der Sloot, B. (Eds.). (2017). *Group privacy: New challenges of data technologies*. Springer. https://doi.org/10.1007/978-3-319-46608-8

Tsesis, A. (2014). The right to erasure: Privacy, data brokers, and the indefinite retention of data. *Wake Forest Law Review*, *49*, 433–484.

Twetman, H., & Bergmanis-Korats, G. (2021). *Data brokers and security. Risks and vulnerabilities related to commercially available data* [Report]. NATO Strategic Communications Centre of Excellence. https://stratcomcoe.org/publications/data-brokers-and-security/17

UK Department for International Trade. (2021, October 22). *G7 Trade Ministers' digital trade principles* [Press release]. https://www.gov.uk/government/news/g7-trade-ministers-digital-trade-principles

United States Senate Committee on Commerce, Science, and Transportation. (2013). *What information do data brokers have on consumers, and how they use it? Hearing before the U.S. Senate Committee on Commerce, Science, and Transportation, 113th Congress*. https://www.govinfo.gov/content/pkg/CHRG-113shrg95838/pdf/CHRG-113shrg95838.pdf

Venkatadri, G., Sapiezynski, P., Redmiles, E. M., Mislove, A., Goga, O., Mazurek, M., & Gummadi, K. P. (2019). Auditing offline data brokers via Facebook's advertising platform. *The World Wide Web Conference on - WWW '19*, 1920–1930. https://doi.org/10.1145/3308558.3313666

Viljoen, S. (2020, October 16). *Data as property? On the problems of propertarian and dignitarian approaches to data governance*. Phenomenal World. https://phenomenalworld.org/analysis/data-as-property.

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, *2019*(2), 494–620. https://doi.org/10.7916/CBLR.V2019I2.3424

Wayne, L. D. (2012). The data-broker threat: Proposing federal legislation to protect post-expungement privacy. *The Journal of Criminal Law and Criminology*, *102*(1), 253–282.

West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, *58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Williams, H. J., & Blum, I. (2018). *Defining second generation open source intelligence (OSINT) for the defense enterprise* [Research report]. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1964.html

World Bank. (2021). *World development report 2021: Data for better lives* [Report]. The World Bank

Group. https://www.worldbank.org/en/publication/wdr2021

Yeh, C.-L. (2018). Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers. *Telecommunications Policy*, *42*(4), 282–292. https://doi.org/10.1016/j.telpol.2017.12.001

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Published by

Alexander von Humboldt Institute for Internet and Society

in cooperation with

CREATe

cnrs centre internet et societe

R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya

University of Tartu Johan Skytte Institute of Political Studies