17. Situating the Marketization of Data

Anne Helmond and Fernando van der Vlist

Abstract

Data are neither inherently valuable, nor do all data have the same value. This contribution argues how data are *made* useful and valuable to specific actors and for specific purposes. It draws attention to the material politics of data flows and valuation, and to the many different actors and stakeholders who build the technological conduits and pipelines that facilitate the circulation and use of data. Therefore, it highlights the need to study the *infrastructural layer* of the global data market, as well as the central role of *intermediaries* who build and uphold these infrastructures for the exchange and use of data for different purposes. Both are important to situate the processes of datafication and data marketization in specific empirical settings.

Keywords: data markets, data intermediaries, marketization, platform infrastructure, partnerships, digital marketing and advertising

Many critical media researchers, technology journalists, and activists have warned about the potential risks and harms of data aggregation and abuses of data by "bad actors," companies, law enforcement agencies, or states. While social media platforms, mobile apps, advertising companies, and data brokers emphasize that the pseudonymous data they collect cannot be traced back to real persons, recent cases have shown how easy it still is to do this.

VICE reported that location data and mobile device data were purchased from a commercial data broker to track the locations of a priest, outing him as gay through his use of the gay/bi/trans/queer dating app Grindr (Cox 2021b). Similarly, government agencies in the US have been known to buy mobile location data from commercial data brokers without warrants for various law enforcement purposes (Guariglia 2020). This commercially available

Es, K. van & N. Verhoeff (eds.), *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023
DOI 10.5117/9789463722971_CH17

data contains unique identifiers, such as mobile device or advertising IDs, that cannot be linked to individuals directly. However, there are many actors in the global data market who offer tools, products, and services to help link and de-anonymize this type of data. This has led to an enormous industry of companies that connect pseudonymous identifiers to a wealth of information obtained from disparate sources, including people's real names, e-mail and home addresses, phone numbers, or credit data. When linked, this information can be used to identify and target individuals or groups of people, thus "shattering" their anonymity (Cox 2021a).

While many of these so-called "data brokers" operate in the shadows, we have learned a lot about them over the years from the critical investigations of many researchers, journalists, and activists (e.g., Beer 2018; Braun 2013; Crain 2021; Christl and Spiekermann, 2016; Lechardoy et al. 2020; Mellet and Beauvisage 2020; Nadler et al. 2018; Zuboff 2019). Additionally, we have done empirical and historical research ourselves into the role of business partners and software infrastructure development in the data economy, which we summarize below (Helmond, Nieborg and van der Vlist 2019; van der Vlist and Helmond 2021; van der Vlist et al. 2022). This research has surfaced some of the key actors, techniques, and technological systems, as well as the material conditions and relations of data as they fuel the advertising-based business models; data-driven business operations; and AI-based tools, products, and services of the contemporary internet. We have shown how the collection, processing, circulation, and use of data impact power relations and raise issues and concerns around the critical political economy of data and data flows.

We have been devising ways of situating data not only in terms of their production contexts but also in terms of their subsequent aggregation, processing, circulation, and use by many different types of users—and often for purposes other than originally intended. Despite what many believe, data are neither inherently valuable nor does all data have the same value. Instead, data are *made* useful and become valuable to specific actors and for specific purposes. Therefore, we draw attention to the materiality and politics of data flows and data valuation and to the many different (intermediary) actors and stakeholders who build the technological conduits and pipelines—or infrastructures—that facilitate the circulation and use of data. By situating the marketization of data in terms of the constitutive actors and infrastructures, we can thus put the opaque global data market in place and in context.

Firstly, our approach highlights the *infrastructural layer* of the global data market. These infrastructures for the exchange and use of data are built by

developers who use application programming interfaces (APIs) to develop data integrations and software applications "on top" of digital platforms. In the global data market, API-based connections between software systems function as the pipelines that enable the circulation and use of data and services between different software platforms and companies. These conduits, once they are built, give other companies and partners the ability to connect, control, and activate data in their own tools, products, and services. We have shown in a large-scale empirical study how this technological infrastructure of API-based integrations between thousands of companies worldwide both provides and *governs* the material conduits for contemporary "programmatic advertising," a multibillion-dollar market that relies on the global data market. With this infrastructure in place, ads and audience commodities are automatically ("programmatically") traded on ad exchanges and served across many media distribution channels and geographic regions in mere milliseconds through real-time bidding auctions. This whole process unfolds in the background each time a consumer opens a web page or uses an app. However, this digital advertising infrastructure also comes with serious risks and harms to society and can be "weaponized by political and anti-democratic actors" to influence political decisions (Nadler et al. 2018), to discriminate, or to otherwise violate people's digital rights (e.g., EDRi 2021).

By identifying who has integrated with, or has built on top of a platform's APIs, we can trace the channels that exist to circulate and use data. Many of these channels are interlinked to enable automated uses and exchanges of data on a large scale, across countries and continents, including through tracking and targeted advertising, and remain an opaque infrastructure for most consumers.

Furthermore, by closely investigating how APIs are designed and structured, we have examined in detail how digital platforms datafy people and their activities as data entities that can be identified and targeted through their associated data fields (e.g., name, birthday, home address, interests, etc.) and connections (e.g., friends, groups, likes, videos, etc.). We traced how data entities such as the "user" have changed and evolved over the years and discovered that Facebook removed sensitive data fields because of ongoing social and regulatory pressures from civil rights organizations and journalists. After Facebook removed data fields related to a user's dating preferences, relationship status, political and religious interests, and friend lists from its Graph API, we found that it kept these data fields available in its Marketing API for advertising and marketing developers for many more years.

APIs not only enable third parties to build applications and services but also provide a powerful means of "infrastructural control" for platform owners to govern, with increasing precision, who is and is not allowed to access data and under which requirements. For example, we observed how Facebook's popular Graph API evolved from a simple interface for data retrieval in the mid-2000s into an increasingly complex and layered "governance arrangement" of (data) access controls, application permissions, app review guidelines, and terms and policies. Additionally, while some "open" APIs are openly available to everyone, the APIs required for digital marketing and advertising are typically governed through special partner programs. Only selected and approved business partners are allowed to access or use platforms' data or to integrate with a platform's technological infrastructure for business purposes. This partnership strategy has been vital for platforms to be embedded in markets and industries other than their own and has led to a complex global data market comprising many interconnected actors and infrastructures.

Secondly, our approach highlights the central role of *intermediaries*, or those who build and uphold these infrastructures of data and automation for different purposes. These are the actors and stakeholders who are doing the practical work of connecting, aggregating, and modeling data from multiple sources (e.g., social media, mobile devices and apps, etc.) and make them available for further uses and users (Beer 2018). Data brokers, data marketplaces, data analytics companies, advertising networks and exchanges, and data management platforms (which enrich advertising bids with tracking data) are all examples of intermediaries in the global data market. Many of these intermediaries are also key "nodes" in the conduits of the global advertising market, enabling others to use their data for digital marketing and advertising purposes. It is the intersection of the global advertising market and the data market that enabled the "inevitable weaponization" of location and app data from Grindr (Cox 2021b).

Additionally, it is important to study these intermediaries to learn how they actually make data useful and valuable. Research on data "assetization" reminds us how it is not the data themselves, but rather the "users" and their "engagement" that are turned into assets. As Birch, Cochrane, and Ward observe, "'users', 'usage', and 'access to users' end up as the legible techno-economic objects that Big Tech can value as future revenue streams through different monetization strategies" (2021, 11; cf. van Doorn and Badger 2020). When advertising online, data generally do not leave "data silos"

like Facebook or Google (as they are called in the industry); rather, these platforms provide interfaces that give customers "access to users."

Popular "identity resolution" services from data partners such as Salesforce, LiveRamp, FullContact, Lotame, and many other companies have become key intermediaries in the global data market. They not only connect and aggregate (audience) data from multiple sources but crucially also make these data available for further uses and users "across the ecosystem." These services typically enrich data sources with additional attributes, such as email addresses, mobile advertising IDs, postal addresses, phone numbers, online or offline purchases, or voting data, to enrich or verify persistent profiles for real persons. Indeed, these services also enabled the de-anonymization of purchased location data that ousted the gay priest using Grindr.

It is common practice for digital platforms or apps to share user data with third parties, including advertising partners, service partners, and social media partners. These data-sharing practices are typically documented in privacy policies. Grindr, for instance, shares device IDs, advertising IDs, and location data with its advertising partners (Grindr 2022). Many of these advertising partners in turn also state in their policies that they share data with third parties, including with their own partners. In short, these partnership strategies are critical to the global data market's functioning and risks. While Grindr's and other companies' policies state that no personally identifiable account information is shared, we know that "identity resolution" services may be used to piece the information together nonetheless and render it personal data, subject to strict regulations like the GDPR.

Even though many of the described data aggregation and sharing practices are forbidden without a user's consent under recent consumer and privacy laws, especially the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), there are plenty of examples of advertising technology companies who have been breaching these legislations as well as complaints from privacy organizations about a lack of and slowness of enforcement efforts by regulatory bodies (Burgess 2022; ICCL 2021; Lomas 2022). The critical industry practice of real-time auction bidding has been found to violate the GDPR because of the industry's inability to trace personal data "behind the scenes" when it passes through the invisible infrastructures of the data market (Ryan and Santos 2022; Veale and Zuiderveen Borgesius 2022). Meanwhile, Apple and Google have begun deprecating the use of third-party cookies and mobile device identifiers (i.e., Identifier for Advertisers [IDFA) on iOS and advertising ID [AdID] on

Android) in the name of consumer privacy. This process is directly impacting the larger ecosystems of actors and technologies relying on Apple and Google's platforms. It has also increased the use of "first-party data" and identity resolution services and has led to the creation of new and competing types of identifiers in the industry (van der Vlist and Helmond 2021).

Taken together, critical perspectives on the technological infrastructures and intermediaries of the global data market enable critical empirical contributions that help us understand the many roles, risks, and harms of data in society. It offers new ways of situating the processes of datafication and marketization in empirical settings. Furthermore, it provides important insights and evidence to help stakeholders, policymakers, and regulators worldwide grapple with the challenges of governing data markets.

References

Beer, David. 2018. "Envisioning the Power of Data Analytics." *Information, Communication & Society* 21, no. 3: 465–79. https://doi.org/10.1080/1369118X.2017.1289232.

Birch, Kean, D.T. Cochrane, and Callum Ward. 2021. "Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech." *Big Data & Society* 8, no. 1: 1–15. https://doi.org/10.1177/20539517211017308.

Braun, Joshua. 2013. "Transparent Intermediaries: Building the Infrastructures of Connected Viewing." In *Connected Viewing: Selling, Streaming, and Sharing Media in the Digital Age*, edited by Jennifer Holt and Kevin Sanson, 134–53. New York: Routledge. https://doi.org/10.4324/9780203067994.

Burgess, Matt. 2022. "How GDPR Is Failing." *WIRED*, May 23, 2022. https://www.wired.com/story/gdpr-2022.

Christl, Wolfie, and Sarah Spiekermann. 2016. *Networks of Control: A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*. Vienna: Facultas Universitätsverlag.

Cox, Joseph. 2021a. "Inside the Industry That Unmasks People at Scale." *Motherboard* (*VICE*), July 14, 2021. https://www.vice.com/en/article/epnmvz/industry-unmasks-at-scale-maid-to-pii.

Cox, Joseph. 2021b. "The Inevitable Weaponization of App Data is Here." *Motherboard* (*VICE*), July 21, 2021. https://www.vice.com/en/article/pkbxp8/grindr-location-data-priest-weaponization-app.

Crain, Matthew. 2021. *Profit Over Privacy: How Surveillance Advertising Conquered the Internet*. Minneapolis, MN: University of Minnesota Press.

EDRi (European Digital Rights). 2021. "How Online Ads Discriminate." https://edri.org/our-work/how-online-ads-discriminate.

Grindr. 2022. "Privacy and Cookie Policy." https://www.grindr.com/privacy-policy/third-parties.

Guariglia, Matthew. 2020. "Law Enforcement Purchasing Commercially-Available Geolocation Data is Unconstitutional." *Electronic Frontier Foundation*, December 3, 2020. https://www.eff.org/nl/deeplinks/2020/12/law-enforcement-purchasing-commercially-available-geolocation-data.

Helmond, Anne, David B. Nieborg, and Fernando van der Vlist. 2019. "Facebook's Evolution: Development of a Platform-As-Infrastructure." *Internet Histories: Digital Technology, Culture and Society* 3, no. 2: 123–46. https://doi.org/10.1080/24701475.2019.1593667.

ICCL (Irish Council for Civil Liberties). 2021. "Europe's Enforcement Paralysis: ICCL's 2021 GDPR Report." https://www.iccl.ie/digital-data/2021-gdpr-report.

Lechardoy, L., A. Sokolyanskaya, and F. Lupiáñez-Villanueva. 2020. "Transparency in the Business-to Business Commercial Relations in the Online Advertising Market." *Observatory on the Online Platform Economy* 3 (December 25, 2020). https://platformobservatory.eu/news/analytical-paper-transparency-in-the-business-to-business-commercial-relations-in-the-online-advertising-market.

Lomas, Natasha. 2022. "Behavioral Ad Industry Gets Hard Reform Deadline after IAB's TCF Found to Breach Europe's GDPR." *TechCrunch*, February 2, 2022. https://social.techcrunch.com/2022/02/02/iab-tcf-gdpr-breaches.

Mellet, Kevin, and Thomas Beauvisage. 2020. "Cookie Monsters: Anatomy of a Digital Market Infrastructure." *Consumption Markets and Culture* 23, no. 2: 110–29. https://doi.org/10.1080/10253866.2019.1661246.

Nadler, Anthony, Matthew Crain, and Joan Donovan. 2018. "Weaponizing the Digital Influence Machine." *Data & Society*, October 17, 2018. New York: Data and Society Research Institute. https://datasociety.net/library/weaponizing-the-digital-influence-machine.

Ryan, Johnny, and Cristiana Santos. 2022. "An Unending Data Breach Immune to Audit? Can the TCF and RTB Be Reconciled with the GDPR?" SSRN Scholarly Paper, March 23, 2022. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4064729.

van der Vlist, Fernando N., and Anne Helmond. 2021. "How Partners Mediate Platform Power: Mapping Business and Data Partnerships in the Social Media Ecosystem." *Big Data & Society* 8, no. 1: 1–16. https://doi.org/10.1177/20539517211025061.

van der Vlist, Fernando N., Anne Helmond, Marcus Burkhardt, and Tatjana Seitz. 2022. "API Governance: The Case of Facebook's Evolution." *Social Media + Society* 8, no. 2: 1–24. https://doi.org/10.1177/20563051221086228.

van Doorn, Niels, and Adam Badger. 2020. "Platform Capitalism's Hidden Abode: Producing Data Assets in the Gig Economy." *Antipode* 52, no. 5: 1475–95. https://doi.org/10.1111/anti.12641.

Veale, Micheal, and Frederik Zuiderveen Borgesius. 2022. "Adtech and Real-Time Bidding under European Data Protection Law." *German Law Journal* 23, no. 2: 226–56. https://doi.org/10.1017/glj.2022.18.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

About the Authors

Anne Helmond is Associate Professor of Media, Data and Society at Utrecht University. She is part of the focus area "Governing the Digital Society," where she examines the processes of platformization, algorithmization, and datafication from an empirical and historical perspective.
> a.helmond@uu.nl

Fernando van der Vlist is a postdoctoral researcher of media and digital society at Utrecht University and with the Collaborative Research Centre "Media of Cooperation" at the University of Siegen (funded by the German Research Foundation, DFG). He is part of Utrecht University's focus area "Governing the Digital Society" and teaches at the Media and Culture Studies department.
> f.n.vandervlist@uu.nl