



**HAL**  
open science

# Regulatory Spillovers and Data Governance: Evidence from the GDPR

Christian Peukert, Stefan Bechtold, Michail Batikas, Tobias Kretschmer

► **To cite this version:**

Christian Peukert, Stefan Bechtold, Michail Batikas, Tobias Kretschmer. Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science*, 2022, 41 (4), pp.318-340. 10.1287/mksc.2021.1339 . hal-03868320

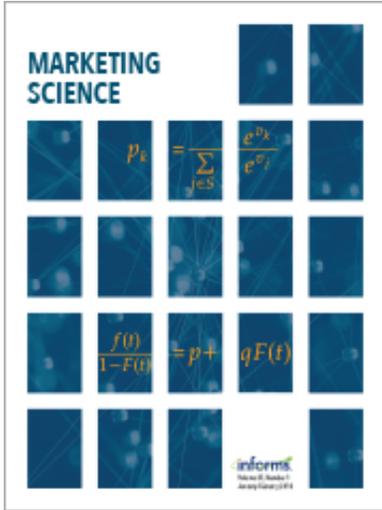
**HAL Id: hal-03868320**

**<https://rennes-sb.hal.science/hal-03868320>**

Submitted on 23 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Regulatory Spillovers and Data Governance: Evidence from the GDPR

Christian Peukert, Stefan Bechtold, Michail Batikas, Tobias Kretschmer

To cite this article:

Christian Peukert, Stefan Bechtold, Michail Batikas, Tobias Kretschmer (2022) Regulatory Spillovers and Data Governance: Evidence from the GDPR. Marketing Science 41(4):318-340. <https://doi.org/10.1287/mksc.2021.1339>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2022 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Regulatory Spillovers and Data Governance: Evidence from the GDPR

Christian Peukert,<sup>a,\*</sup> Stefan Bechtold,<sup>b</sup> Michail Batikas,<sup>c</sup> Tobias Kretschmer<sup>d,e</sup>

<sup>a</sup>Department of Strategy, Globalization and Society, HEC Lausanne, Université de Lausanne, 1015 Lausanne, Switzerland; <sup>b</sup>Department of Humanities, Social and Political Sciences, ETH Zürich, 8092 Zurich, Switzerland; <sup>c</sup>Supply Chain Management and Information Systems, Rennes School of Business, 35065 Rennes, France; <sup>d</sup>Institute for Strategy, Technology and Organization, LMU Munich School of Management, 80539 Munich, Germany; <sup>e</sup>CEPR Centre for Economic Policy Research, London EC1V 0DX, United Kingdom

\*Corresponding author

Contact: christian.peukert@unil.ch,  <https://orcid.org/0000-0003-3997-8850> (CP); sbechtold@ethz.ch (SB); michail.batikas@rennes-sb.com,  <https://orcid.org/0000-0003-0833-5086> (MB); t.kretschmer@lmu.de,  <https://orcid.org/0000-0002-7830-7825> (TK)

Received: June 25, 2020

Revised: July 14, 2021; October 6, 2021

Accepted: October 8, 2021

Published Online in Articles in Advance:  
February 15, 2022

<https://doi.org/10.1287/mksc.2021.1339>

Copyright: © 2022 The Author(s)

**Abstract.** We document short-run changes in websites and the web technology industry with the introduction of the European General Data Protection Regulation (GDPR). We follow more than 110,000 websites and their third-party HTTP requests for 12 months before and 6 months after the GDPR became effective and show that websites substantially reduced their interactions with web technology providers. Importantly, this also holds for websites not legally bound by the GDPR. These changes are especially pronounced among less popular websites and regarding the collection of personal data. We document an increase in market concentration in web technology services after the introduction of the GDPR: Although all firms suffer losses, the largest vendor—Google—loses relatively less and significantly increases market share in important markets such as advertising and analytics. Our findings contribute to the discussion on how regulating privacy, artificial intelligence and other areas of data governance relate to data minimization, regulatory competition, and market structure.

**History:** Avi Goldfarb served as the senior editor and Anja Lambrecht served as associate editor for this article.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “Marketing Science. Copyright © 2022 The Author(s). <https://doi.org/10.1287/mksc.2021.1339>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

**Funding:** This work was supported by the Portuguese Foundation of Science and Technology [Grant UID/GES/00407/2013]. C. Peukert acknowledges support from the Swiss National Science Foundation [Project 100013\_197807]. T. Kretschmer acknowledges support from the German Research Foundation (DfG) [Project KR 3479/8-1].

**Supplemental Material:** The online appendix and data are available at <https://doi.org/10.1287/mksc.2021.1339>.

**Keywords:** privacy • competition policy • antitrust • internet regulation • regulatory competition • compliance risk • GDPR • Brussels effect • cookies • web tracking • data governance • data minimization

## 1. Introduction

The Internet has revolutionized global trade and data flows. In many aspects of our daily life, it has torn down national borders and facilitated communication and trade across the globe. Regulating such a world is challenging. As international coordination mechanisms have often proven ineffective, individual countries and regions have enacted legal regimes to govern the data-driven world. In the area of privacy, examples of such regimes include the European Union (EU)’s General Data Protection Regulation of 2018 (GDPR) and the California Consumer Privacy Act of 2020 (CCPA).

Privacy protection in Europe has traditionally been strong for historical, cultural, political, and legal reasons (Bradford 2020, p. 136–141; Schwartz and Peifer 2017, p. 123–127). The GDPR is the cornerstone of

European privacy law and is considered the most comprehensive, globally leading privacy regime. It establishes common rules on data processing throughout the EU and is directly binding for companies and residents in the EU and beyond. With the GDPR, the European legislator intended to harmonize privacy law and enforcement throughout the EU and increase the protection of individuals’ privacy while maintaining the benefits of data processing.

We investigate empirically whether and how the way websites, web technology providers, and consumers interact has changed with the GDPR both within and outside the EU and explore changes in the structure of markets for web technologies. Websites may use web technologies to raise advertising revenues, observe user behavior, share information through social media, or

host audiovisual content. We observe whether a website uses such technologies through the HTTP requests the website makes to external servers and map these requests to third-party firms. We also collect the stated privacy policies of these vendors. We follow more than 110,000 websites from May 2017 to November 2018. With data covering 12 months before and 6 months after the GDPR came into force, we can document the short-run changes in how websites interact with web technology providers, as well as changes in the web technology market in the same time frame.

We highlight three key findings. First, websites have reduced the number of third-party domains they request after the GDPR became effective. Importantly, these changes also apply to websites which are not legally subject to the GDPR. Although the GDPR may apply to firms in the EU catering to consumers outside the EU and to firms outside the EU catering to consumers in the EU, it does not, *de jure*, apply to firms outside the EU catering to consumers outside the EU. Nevertheless, we observe a *de facto* change by websites located outside the EU serving consumers outside the EU in our empirical setting. This hints at a “Brussels effect,” which has long been described in the legal literature on regulatory competition and the EU (Bradford, 2012, 2020): Our findings suggest that the EU has shaped the global privacy regime beyond the boundaries of the EU, thus exporting the EU’s regulatory framework to other countries.

Second, the number of third-party domains requested by websites is reduced immediately after the GDPR becomes effective, but the decades-long general upward trend remains unchanged. However, focusing on a specific type of request—cookies—we find that websites’ use of third-party technologies that interact with consumers’ privacy has changed effectively with the GDPR: There is a sustained decrease in third-party cookies after the introduction of the GDPR. The change in the interactions between websites and third-party vendors is especially pronounced for vendors disclosing that they collect personal data. Together with the finding that the stated privacy policies of technology vendors become more informative about collection, processing and sharing of personal data, this seems in line with the GDPR’s data minimization principle.

Third, the market for web technologies has changed substantially after the introduction of the GDPR. Although the market overall is shrinking as websites send less requests to third-party vendors, the dominant firm in many markets for web technologies, Google, increases its market share. This indicates that privacy regulation may have unintended consequences for market structure and competition.

We make several contributions. First, by carefully analyzing the institutional arrangement of the GDPR, we disentangle its implications for firms and consumers

located within or outside the European Union. Second, using a unique data set with information on the location of both firms and consumers, we observe how various geographical markets evolve differently after the GDPR becomes effective. Third, we study changes to pre-existing trends following the GDPR. With the caveat of a relatively short observation period of six months after the GDPR, we can speculate about the longer-run impact of the GDPR when comparing immediate changes (intercept) versus changes in the growth rate (slope). Fourth, using data on the privacy policies of web technology vendors, we document changes in the disclosure of collection and sharing of personal data. Finally, we find that privacy law interacts with other policy dimensions. Even websites outside the legal scope of the GDPR at least temporarily change operations, and the market structure of web technology vendors changes. Hence, we add to the discussion on how regulating privacy and other areas of data governance, for example, artificial intelligence, relates to data minimization, regulatory competition, and market structure.

## 2. Institutional Background and Related Literature

With the GDPR, European legislators intended to harmonize privacy law and enforcement throughout the EU and increase the protection of individuals’ privacy while maintaining the benefits of data processing. The GDPR became applicable on May 25, 2018 (European Union 2016) and is binding for firms and residents in the EU and beyond. Three aspects of the GDPR are especially relevant for our paper: territorial application, increased compliance risks, consent and data minimization.

### 2.1. Territorial Application

Any empirical study on the evolution of various geographical web markets after the GDPR became effective faces the challenge of the territorial applicability of the GDPR. Countries are typically entitled to enact laws and exercise authority within their geographical boundaries. The Internet has challenged the concept of geography-based rule setting. Users in one country can browse, communicate, and shop on any website located anywhere around the globe, calling into question where the service takes place. In the context of the GDPR, this leads to situations in which users located within or outside the EU access websites located within or outside the EU. In theory, the GDPR clearly regulates in which of these cases the regulation is applicable or not. As Table 1 illustrates, the GDPR applies, first, in the standard case in which both the user accessing a website and the website processing the user’s personal data are located in the EU (Art. 3(1) GDPR, top left cell in Table 1). We will refer to this as Case 1. In this context, the user’s citizenship, residence, or legal status is irrelevant

**Table 1.** Territorial Scope of the GDPR

		Website location	
		EU	Non-EU
Website audience	EU	Case 1: Standard case GDPR applies Art. 3(1) GDPR	Case 3: Effects doctrine GDPR applies Art. 3(2) GDPR
	non-EU	Case 2: Establishment principle GDPR applies Art. 3(1) GDPR	Case 4: “Brussels effect” GDPR does not apply

*Note.* Table shows whether the GDPR is applicable, as a matter of law, to Internet users and firms located within and outside the European Union.

(European Data Protection Board 2019a, p. 14–16); only the user’s location matters. Second, the GDPR applies if users located outside the EU access a website in the EU that processes their personal data (Case 2, following the so-called “establishment principle,” Art. 3(1) GDPR, bottom left cell in Table 1; European Data Protection Board 2019a, p. 8–13). Third, the GDPR applies if users in the EU access a website located outside the EU which processes their personal data (Case 3, so-called “effects doctrine,” Art. 3(2) GDPR, top right cell in Table 1). It is sufficient that a firm intentionally offers goods or services to people located in the EU, or if it monitors or predicts behavior, personal preferences, or attitudes within the EU (Art. 3(2) and Recital 24 GDPR; European Data Protection Board 2019a, p. 15–20). Whether a firm targets persons located in the EU is decided on a case-by-case basis. This involves, *inter alia*, assessing the website’s language, the currency used, and the firm’s marketing efforts (Recital 23 GDPR).

As the discussion of these three situations illustrates, the EU has designed the GDPR to be widely applicable to a range of activities within and outside the territorial boundaries of the EU. The GDPR can effectively apply to websites and web technology providers regardless of their business location or legal incorporation as long as they are accessed by and used for users in the EU. Websites and web technology providers not located in the EU but subject to the GDPR must designate a representative in the EU (Art. 27 GDPR). They are bound by all rules of the GDPR, including its damages and fines regime. Compared with former European privacy law, this is a drastic expansion of the global reach of European privacy protection. Under Art. 4(1)(c) of the former Data Protection Directive (European Union 1995), EU privacy law only applied if the data controller used equipment located in the EU, applying a territoriality doctrine. The GDPR broadened its territorial scope by switching to an effects-based doctrine.

The only situation in which the GDPR does *not* legally apply is if users located outside the EU access a website located outside the EU that processes their personal data (Case 4, bottom right cell in Table 1). Although *de jure*, the GDPR does not apply in this situation, the

proponents of the so-called Brussels effect argue that the GDPR *de facto* does apply in this situation as well.<sup>1</sup> Our analysis in Section 5.1.1 can be seen as an empirical test whether such a Brussels effect exists with regard to the GDPR: Do websites which are not required by law to obey the GDPR follow GDPR rules nevertheless?

Defining the boundaries between the four cases in Table 1 can be a challenging endeavor in its own right. As it may take years to resolve issues of territorial application questions,<sup>2</sup> the GDPR left firms outside the EU with considerable legal uncertainty during our period of empirical observation whether and to what extent the GDPR applied to them.

## 2.2. Increased Compliance Risks

The GDPR has drastically increased compliance risks for privacy violations for several reasons. First, under pre-GDPR European privacy law, maximum fines for privacy violations varied between €12,000 and €600,000 and were set at the EU member state level with considerable heterogeneity in enforcement. Fines are now up to €20 million or 4% of the total worldwide annual turnover, whichever is higher (Art. 83 (5) and (6) GDPR). European privacy law now resembles European antitrust law, where the European Commission has issued several multibillion dollar fines over the last decade. The GDPR also initiated far-reaching changes for the competence and cooperation of national data protection authorities responsible for monitoring GDPR compliance.

Second, under the GDPR, a website cannot easily dispose of its liability for privacy violations by outsourcing the processing of personal data—for example, for behavioral monitoring or consumer profiling—to third-party web technology providers. This is not only because the website must inform its users about any transfer of personal data to third-party providers (Art. 13 (1)(e) GDPR) and because the website must make sure that the third-party providers’ data processing will adhere to the GDPR as well (Art. 28 (1) GDPR). The website may also be jointly responsible with the web technology provider for violating provisions of the GDPR.<sup>3</sup> This legal framework of joint responsibilities can have important implications for websites. For example, it is the websites and

not just Google or Facebook which have to gather user consent to process personal data. Also, users can sue a website for damages caused by GDPR violations committed by its web technology provider within the joint responsibility framework.<sup>4</sup> Although websites may be able to outsource tasks such as analytics to third-party providers, they are—to a considerable extent—still responsible and potentially liable for privacy violations that occur while cooperating with the technology provider.

Third, the GDPR has put compliance with privacy laws high up on firms' agendas not only because of its stricter enforcement regime and broader territorial scope, but also because the GDPR has led to novel problems in interpreting the GDPR and its relationship to other bodies of EU and member state privacy laws. For example, during the period of our study, it was difficult to observe a standard pattern on how European data protection authorities calculated fines, how the territorial scope of the GDPR was determined in individual cases (Tobin 2019), or whether a non-EU website "targeted" users located in the EU (see Section 2.1).

Such questions can only be resolved by data protection authorities providing guidelines on how to interpret the GDPR—which may take considerable time<sup>5</sup>—and ultimately by courts deciding disputes. As this process typically takes several years at least, firms were left with increased legal uncertainty and considerable compliance risks after the GDPR was introduced (Bessen et al. 2020).<sup>6</sup>

### 2.3. Consent and Data Minimization

If a website uses web technologies such as cookies,<sup>7</sup> scripts, images, or fingerprinting technologies to identify users, the website complies with the GDPR if it gathers the consent of the users whose personal data it processes (Art. 6(1)(a) GDPR).<sup>8</sup> For cookies, the revised e-Privacy Directive (European Union 2002) had mandated websites to gather user consent for placing cookies on their devices since 2009 (Article 5 (3) of the revised e-Privacy Directive).<sup>9</sup> The GDPR goes much further as it covers the processing of any kind of personal data, including but not exclusively through cookies. In addition, the drastic expansion of the sanctions regime under the GDPR and the broad expansion of the territorial reach of the GDPR acted as a game-changer. With its wide-ranging, technology-neutral rule on consumer consent, the GDPR gave European privacy authorities a powerful tool to enforce consumer consent across the EU and beyond.

Furthermore, the GDPR implements a data minimization principle (Art. 5(1)(c), 25(1) and Recitals 78, 156 GDPR), where personal data collection needs to be limited to what is necessary for legitimate processing purposes. Although it is difficult to measure data minimization in general, our empirical context lets us study some aspects consistent with data minimization

in the sense of the GDPR. We observe whether websites engage in less interactions with third-party technology providers after the introduction of the GDPR. Such interactions can, from a technical perspective, be considered means to collect and share personal data. Whether a relative decline implies *data minimization* is impossible to establish, but it would at least be informative of *data reduction*.

### 2.4. Related Literature

We add to an emerging empirical literature on the technical and commercial implications of the GDPR. Related to our study, Johnson et al. (2020) also report a decrease of third-party requests by 15% and an increase in the concentration in the industry of web technologies by 17% after the introduction of the GDPR. In contrast to our work, they use a shorter pre-GDPR period and only follow the top websites per country. As we show, the change in interactions with third-party web technology vendors is more pronounced in the tail than the top of the popularity distribution, suggesting that the GDPR has disproportionately affected smaller organizations. Moreover, our rich data set yields additional insights on changes in the web technology industry after the introduction of the GDPR. We show that the increase in concentration can be traced back to an increase in the market share of the dominant vendor. We document that the market as a whole shrinks, that is, the dominant vendor also receives less third-party requests, but as its decline is relatively less pronounced, it takes a bigger piece of a shrinking pie. Our paper therefore provides detailed insights consistent with some of the results in other works, including Johnson et al. (2020), industry reports (WhotracksMe 2018), and recent academic work in computer science and communication (Libert et al. 2018, Dabrowski et al. 2019, Degeling et al. 2019, Hu and Sastry 2019, Solomos et al. 2019, Sørensen and Kosta 2019, Urban et al. 2020). It is worth mentioning that our work also confirms theoretical work from the pre-GDPR era (Gopal et al. 2018), in which the authors show that increased privacy concerns decrease the number of third parties to which websites are connected, resulting in higher concentration in the third party industry. Moreover, our work identifies websites with different regulatory exposure to the GDPR based on target audience and geographical location of the organizations behind websites. This lets us provide important evidence on the extraterritorial reach of the EU's regulation. Finally, we find suggestive evidence that the GDPR's goal of data minimization seems to have been reached: Web technology providers that use cookies are requested less following the GDPR, and web technology providers tend to adopt more transparent privacy policies following the GDPR.

We add to a stream of literature linking the enactment of the GDPR to different outcomes. It has been

reported that the GDPR had a negative impact on publishers' performance (Lefrere et al. 2019, Aridor et al. 2020, Schmitt et al. 2020, Goldberg et al. 2021) and on European firms' ability to attract investment (Jia et al. 2021). In line with these results, Sharma et al. (2019) found that in stricter privacy regimes, small publishers, and small advertisers see their profit decline. Moreover, the GDPR seems to have negatively impacted innovation, because AI startups are reallocating their limited resources to deal with the implications of the GDPR (Bessen et al. 2020). Furthermore, the GDPR has decreased the entry rate of new mobile apps in the market (Janssen et al. 2021). However, some studies find a neutral or even positive effect of the GDPR. Godinho de Matos and Adjerid (2021) show that consumers' opt-in decisions have increased after the GDPR, leading to an increase in sales because of more effective targeted advertising, whereas Zhuo et al. (2021) finds that the GDPR had no impact on Internet interconnection. Finally, theoretical work shows that, although the GDPR can increase consumer surplus, it can negatively affect firm profits, especially in competitive markets (Ke and Sudhir, 2020).

Our study also contributes and relates to theoretical work on the antitrust implications of the GDPR (Jia et al. 2021, Gal and Aviv 2020, Geradin et al. 2021, Economides and Lianos 2021) and on regulatory competition (Bradford 2012, 2020; Goldfarb and Trefler 2018; Frankenreiter 2022). We show that regulating privacy can affect market structure and competition. In this respect, our paper relates to work suggesting that larger firms can disproportionately benefit from data-enabled learning (Farboodi et al. 2019, Hagiu and Wright 2020) and that privacy regulation can increase market concentration by restricting data flows across firms (Campbell et al. 2015, Acemoglu et al. 2019, Jin and Wagman 2020, Jones and Tonetti 2020). This also speaks to the literature on heterogeneous effects of regulation (Elliehausen and Kurtz 1988, Elliehausen 1998, Dahl et al. 2016).

### 3. Web Technologies and Personal Data

Modern websites and web applications tend to be highly modular, and website operators can access a plethora of web technologies provided by third-party vendors. These technologies are usually offered in a software-as-a-service model, making them distinct from the traditional licensing of software libraries or packages that can be installed and run on premise. The fundamental concept is one of outsourcing. Applications running on external hardware are accessed over the Internet and integrated into the own technology stack. Early discussions on the business implications of the underlying technological concepts ("XML Web Services" and "Application Programming Interfaces")

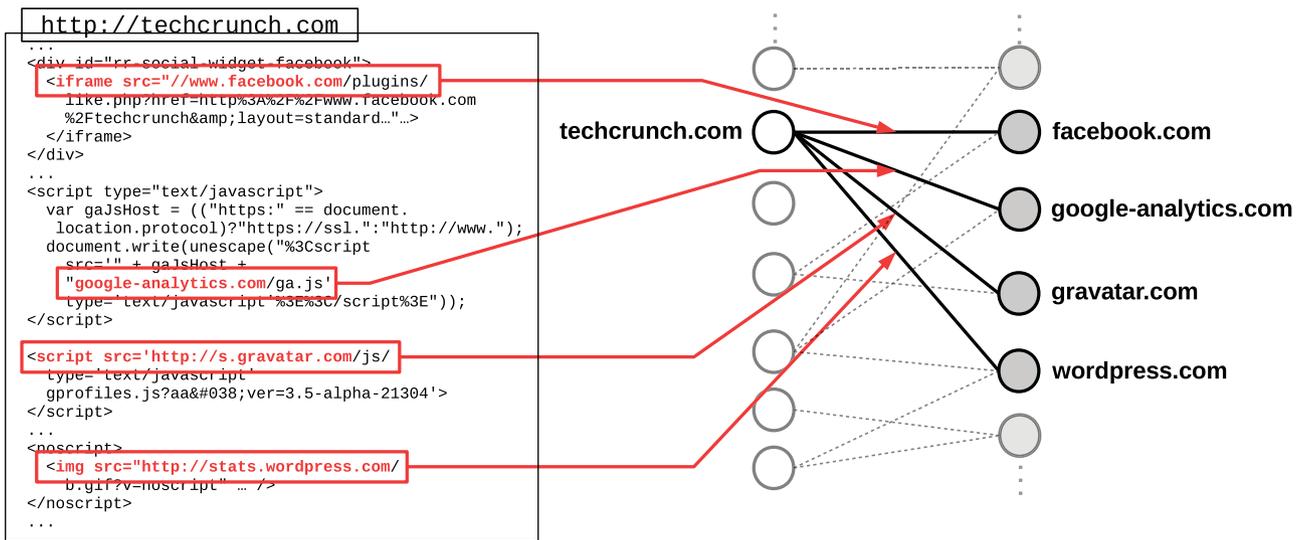
name interoperability and scalability as key characteristics, which at the time added "another dimension to the Web; instead of just person-to-person or person-system, it also handles system-to-system" (Lim and Wen 2003, p. 50). Most importantly for the commercial Internet, these advances enabled the emergence of online advertising as a means to generate website revenues. In general, however, web technologies span a wide variety of functions, serving different technological and business purposes.

First-order choices in the development of websites and web applications include frameworks and libraries where and how functionality and content is hosted, and the monetization model.<sup>10</sup> These choices are correlated with the categories of third-party web technologies a website would adopt. More generally, the reasons for outsourcing rather than relying on in-house technology have been discussed in the large information technology (IT) outsourcing literature (starting with Loh and Venkatraman 1992), including price and relative cost advantages (Ang and Straub 1998), service quality (Grover et al. 1996), and strategic considerations (Watjatrakul 2005) as drivers and success factors of IT outsourcing.

We focus on web technologies that require real-time interaction across entities. These include technologies that optimize content delivery to a particular user, provide personalized content, help publishers and advertisers understand their audience, and services connecting publishers to advertisers. These interactions can be observed via the HTTP requests that a website makes as content is loaded. Using HTTP requests, websites can display content controlled by other parties, and this content can differ for each request and user. Figure 1 illustrates a stylized example of third-party requests: the page hosted on *techcrunch.com* makes requests to third-party domains such as *google-analytics.com* and *wordpress.com*, embedding content hosted on third-party servers, such as scripts and images.

Over the last 20 years, the web technology service industry has grown by a factor of 50 in terms of the number of distinct services on the market and by a factor of 4 in terms of the median number of distinct services used per website (Lerner et al. 2016). About 88% of the top 1 million most popular websites operate with at least one third-party web technology service, and out of those, the average site requests about nine distinct third-party domains (Libert 2015). The long-run trend of an increasing number of third-party requests can be linked to the increasing modularity and complexity of modern web applications (Cheng et al. 2006, Yoo et al. 2012). At the same time, the web technology industry has become considerably concentrated with the top 20 services covering about a third of the market (Gill et al. 2013, Schelter and Kunegis

Figure 1. (Color online) Example of Third-Party Requests



Note. Similar to Figure 1 in Schelker and Kunegis (2018).

2018). Supply-side economies of scale, network effects, and the vertical integration of large vendors into more service categories can reinforce the cost and quality benefits from consolidation (Currie 2000). Furthermore, the broader literature on IT outsourcing suggests that regulation—mostly on IT security and risk management—drives firms to choose larger vendors and bundled service offers for compliance reasons (Willcocks et al. 2010).

The web technology industry is an ideal setting for studying firm responses to privacy regulation. First, historical data on third-party requests of a large number of websites is publicly available and can be matched with meta-information on service categories and privacy policies of vendors. Second, all third-party services called in real-time as a user navigates to the website can be used to collect and share personal data. For example, a service can trace back the IP address from which an HTTP request originated (i.e., the IP address of the user). It can also respond to a request by sending a cookie to be stored on the user's machine, which lets the service recognize the same machine in the future, even if the IP address has changed. The technological possibility— independent of whether specific services actually do collect, process or share personal data—implies compliance risks for website operators interacting with third-party services. We distinguish between third-party requests in general as technologies that could be used to collect, process, and share personal data and third-party requests that respond with cookies. The latter, by definition, collect information that allows identification of a device, and thus (with varying degrees of accuracy; Díaz-Morales 2015) the user of the device.

## 4. Data and Methods

### 4.1. Data

We combine various public and proprietary data sources to study the behavior of websites, web technology providers, and consumers before and after the GDPR became effective. Table 2 provides an overview. We describe the data sources' content and some key descriptive statistics here.<sup>11</sup>

**4.1.1. Main Data Set: HTTPArchive.** Our main data set contains historical information about websites' HTTP requests to third parties. The documentation project *HTTPArchive* periodically crawls the homepages of about half a million hosts. We mostly use data from May 2017 to November 2018, with a longer pre-GDPR period as a robustness check. With a few exceptions, these data are available on a biweekly basis, stemming from a crawl at the beginning and in the middle of the month.<sup>12</sup> The length of our panel lets us capture pre-existing trends and study longer-run dynamics. The level of observation is a "website-host," for example, `subdomain.domain.com`, mostly of the form `www.domain.com`.<sup>13</sup> We construct a balanced sample of hosts present throughout our study period, resulting in 110,706 hosts over 33 points in time—21 data points before the introduction of the GDPR and 12 data points after, resulting in 3,653,298 observations in total.<sup>14</sup>

We collect information about the total number of requests to third-party domains (mean, 15.6),<sup>15</sup> the identity of these domains, and the number of domains that respond to a request by sending a cookie (mean, 6.7). There is substantial variation in these numbers across websites with (non)EU audiences and (non)EU

**Table 2.** Variables and Data Sources

Information	Data set	Type	Uniques	Period
<b>Websites</b>				
<i>Interactions with vendors</i>				
Third-party requests and cookies	HTTPArchive	Public	110,706	05/2017–11/2018
<i>Audience</i>				
Top-level domain	HTTPArchive	Public	110,706	05/2017–11/2018
Language	HTTPArchive	Public	16,963	05/2017–11/2018
Country-specific traffic rank	Alexa	Public	110,706	As of 09/2019
<i>Location</i>				
Country of company behind website	Crunchbase	Proprietary	34,950	Accessed 09/2020
Country of domain registrant	whoisxmlapi.com	Proprietary	100,717	Accessed 10/2020
<i>Categories</i>				
Content categories	Nielsen	Proprietary	30,265	Accessed 07/2017
<i>Technology vendors</i>				
Third-party domains requested	HTTPArchive	Public	63,812	05/2017–11/2018
Technology categories	whotracks.me	Public	1,631	Accessed 02/2019
Company behind domain	whotracks.me	Public	1,890	Accessed 09/2019
Privacy policies	Evidon	Public	5,053	09/2017 and 07/2019

locations and over time—in particular, before and after the GDPR came into force.

**4.1.2. Website Location.** Although our level of analysis is at the website host, the location information can only be measured at the domain level. Hence, we make the assumption that the company/owner behind `subdomain1.domain.com` is identical to the company/owner behind `subdomain2.domain.com`.

**4.1.2.1. Crunchbase.** We obtain data from Crunchbase to access information about the physical location of the headquarters of the company behind the website domain we observe in the *HTTPArchive*. We link *HTTPArchive* to Crunchbase based on the company website listed on a Crunchbase profile. We observe country information from Crunchbase for 34,950 (31.6%) websites.

**4.1.2.2. WHOIS.** We further obtain information about the owner of the website through WHOIS requests in the domain registry database. Using the proprietary service *whoisxmlapi.com*, we can identify the country of the domain registrant for 94,919 (85.7%) websites in our sample.<sup>16</sup>

**4.1.3. Website Audience.**

**4.1.3.1. Country-Specific Top-Level Domains.** We primarily infer whether a website-host caters to an EU audience via the country-specific top-level domain (TLD). EU-specific TLDs include the country-specific domains of the EU member states, as well as `.eu`.

**4.1.3.2. EU-Specific Languages.** An alternative way to infer website audience is to extract meta-information from the website, notably its language and thus its likely target audience. We use

*HTTPArchive* to determine whether websites use one of the 23 official languages in the EU (excluding English).

**4.1.3.3. Country-Specific Traffic.** We add information on country-specific demand from *Alexa*. Specifically, we obtain the rank by country (based on page views, as of October 2019), if available, for every country in the world. These data are incomplete, but we can add this information for 42,764 websites. We define a website as catering to an EU audience if it appears in *Alexa*'s ranking in at least one EU country. The country-specific *Alexa* ranks cover a wide range of the popularity distribution. Websites in our baseline sample have an average minimum rank of one and an average maximum rank of 37,530 in the per-country popularity lists. Regarding the global ranking, all the websites included in the sample, except 99 websites, are part of *Alexa*'s top one million list, with the median ranking being 158,774.

**4.1.4. Definition of Territorial Scope of the GDPR.** We apply any of the previous criteria to define the four different cases of Table 1. Regarding website location, a website is an EU website (i) if its headquarters are in EU (Crunchbase data) or (ii) if the country of the domain registrant is an EU country (WHOIS data). Regarding the audience location, a website serves an EU audience (i) if the TLD of the website is an EU-specific top-level domain or (ii) if the website uses 1 of the 23 official languages of the EU (excluding English) or (iii) if the website has traffic from at least one EU country (*Alexa* data). After applying this definition, we end up with 19.4% of our sample of websites belonging to Case 1 of Table 1, 11.7% to Case 2, 6.8% to Case 3, and finally 61.9% to Case 4. In Table 3, we provide summary statistics regarding the third-party requests and

**Table 3.** Summary Statistics Before and After the GDPR

	N	Before GDPR				After GDPR			
		Mean	Standard deviation	Minimum	Maximum	Mean	Standard deviation	Minimum	Maximum
Standard case (EU–EU)									
Requests	21,527	15.2071	14.00868	0	173.48	15.12444	15.41681	0	142.3333
Cookies	21,527	6.9607	9.891918	0	141.2	6.532516	10.7182	0	99.83334
Effects doctrine (non-EU–EU)									
Requests	12,973	18.44721	16.8271	0	158.68	18.90187	18.31705	0	184.5833
Cookies	12,973	8.602379	12.01794	0	125.48	8.394873	12.74905	0	133.9167
Establishment principle (EU–nonEU)									
Requests	7,600	12.80002	12.49687	0	130.28	12.66103	13.10103	0	125.9167
Cookies	7,600	5.230495	8.92324	0	97.12	4.821261	9.004335	0	87.5
Brussels effect (non-EU–non-EU)									
Requests	68,567	15.25551	14.79574	0	191.12	15.74936	15.31034	0	173.5833
Cookies	68,567	6.537653	10.31434	0	149.56	6.424841	10.23933	0	135.9167

cookies of each one of these website categories before and after the GDPR.

#### 4.1.5. Technology Vendor Information.

**4.1.5.1. Service Categories.** We obtain meta-information about the categories of third-party web technology providers in the *HTTPArchive* from *whotracks.me*. The classification of *whotracks.me* includes 12 narrow categories (Audio/Video Player, Comments, Customer Interaction, Essential, Analytics, Social Media, Misc, CDN, Hosting, and Extensions). We use all of these categories for our analysis of the extent of services (Section 5.3.2) and focus only on the five categories *Audio/Video Player*, *Advertising*, *Analytics*, *CDN*, and *Other* that are most important in terms of market share for Google in our sample. We define *Other* to include the categories *Misc* and *Hosting* from the original *whotracks.me* classification. The results reported below are robust to variations of these definitions.

**4.1.5.2. Mapping Third-Party Domains to Companies.** *whotracks.me* also gives information about the companies behind the third-party domains we observe in *HTTPArchive*. We only have this information for 1,890 domains and assume that all domains that we cannot link to a particular company are independent companies. The average number of websites served by third-party domains that we cannot link to a company is 8.4, whereas the average number of websites served by third-party domains that we can link to a company is 1,247.9.

**4.1.5.3. Privacy Policies.** Finally, we have snapshots of *Evidon's* industry directory that provides information on privacy policies, in particular, whether a web technology provider discloses in its privacy policy whether it collects and shares data with third parties. In these data, we can distinguish between anonymous, pseudonymous, aggregate, personally identifiable, and sensitive

(relating to personal financial or health information) data.

#### 4.2. Econometric Model

The empirical setting of the GDPR makes it challenging to find a valid control group to establish a counterfactual. However, we believe it is possible to treat the introduction of GDPR as an exogenous shock (see Section A.2 in the online appendix for a more detailed discussion). Although the lack of a clear counterfactual makes it difficult to make strong causal claims, we set up a model allowing us to document the changes that narrowly coincide with the timing of the introduction of the GDPR. Hence, after ruling out a few alternative explanations, such as changes in the industry because of tracking prevention technology and privacy scandals, we conclude that unobserved factors that correlate closely with the timing of the introduction of the GDPR are likely negligible. Hence, it seems reasonable to expect that—at least in the short run—our estimates closely resemble the causal effect of the GDPR. Our baseline specification is as follows:

$$\begin{aligned} \text{Log}(D_{it} + 1) = & \beta_1(\text{Trend}_t \times \text{EU-Audience}_i) \\ & + \beta_2(\text{Trend}_t \times \text{nonEU-Audience}_i) \\ & + \gamma_1(\text{Trend}_t \times \text{Post}_t \times \text{EU-Audience}_i) \\ & + \gamma_2(\text{Trend}_t \times \text{Post}_t \times \text{nonEU-Audience}_i) \\ & + \delta_1(\text{Post}_t \times \text{EU-Audience}_i) + \delta_2 \\ & (\text{Post}_t \times \text{nonEU-Audience}_i) + \mu_i + \varepsilon_{it}, \quad (1) \end{aligned}$$

where  $D_{it}$  is the number of third-party domains that website  $i$  requests at time  $t$ . In other specifications, we also look at the subset of third-party domains that respond with at least one cookie.  $\text{Post}_t$  indicates the period after the GDPR came into force on May 25, 2018. We include group- and time period-specific linear time trends and website fixed effects  $\mu_i$ . The error term  $\varepsilon_{it}$  has the standard assumptions, and we report estimates clustered at the website level.

This model lets us carry out a before/after comparison within the four groups of websites in Table 1. In particular, we are interested in estimating a discontinuity around the introduction of the GDPR. That is, we are more interested in estimating the  $\delta$  parameters than the  $\beta$  or  $\gamma$  parameters in Equation (1). Estimating the model on the subsample of websites located in the EU,  $\delta_1$  gives an estimate of the change in  $D_{it}$  for websites located in the EU and catering to EU audiences (Case 1 in Table 1), whereas  $\delta_2$  is the estimate of the change in  $D_{it}$  for websites located in the EU and catering to non-EU audiences (Case 2 in Table 1). Estimating Equation (1) on the subsample of websites not located in the EU,  $\delta_1$  and  $\delta_2$  are the respective estimates of the change in  $D_{it}$  for Cases 3 and 4 in Table 1).

### 5. Findings

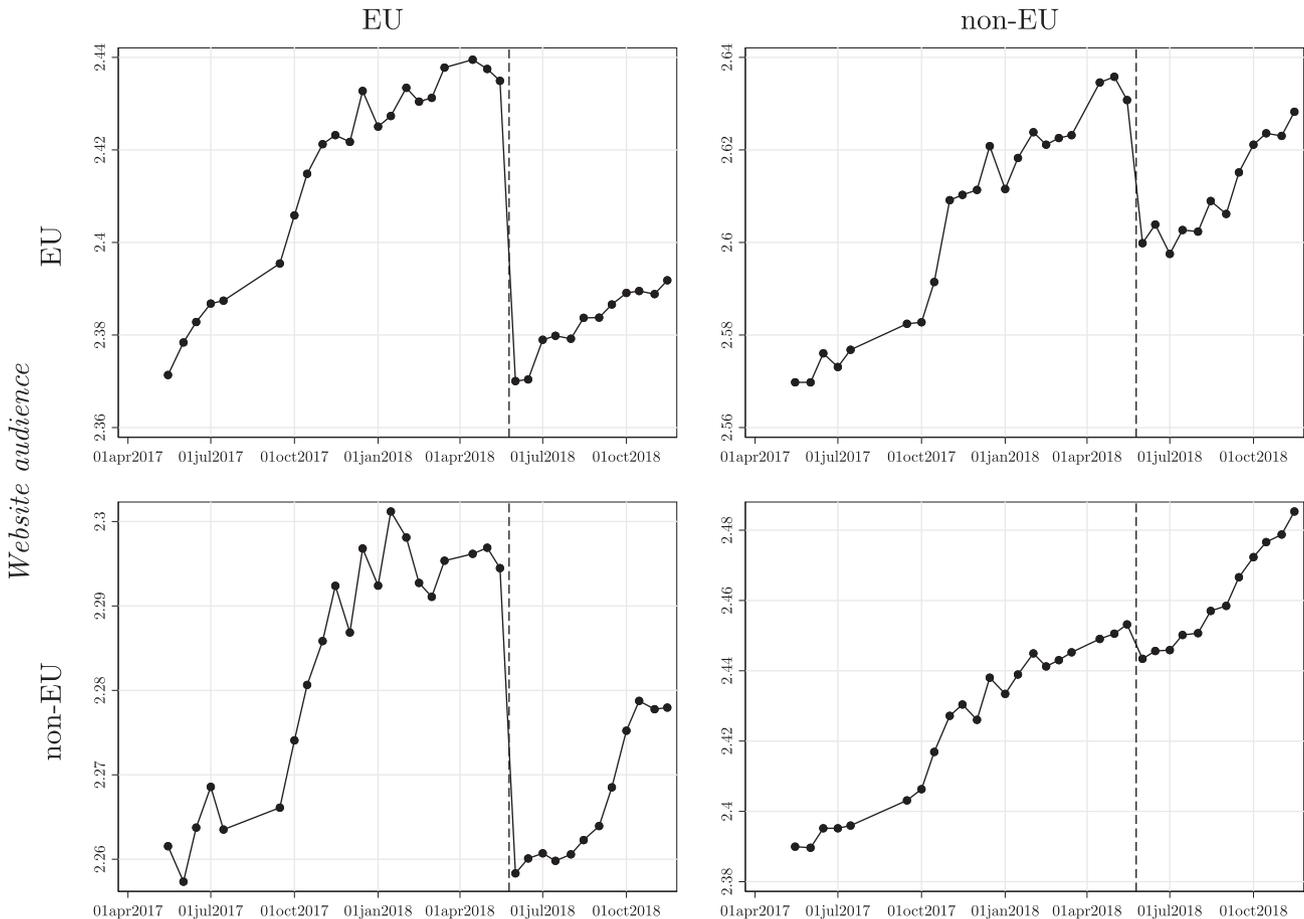
We structure our findings along three dimensions: First, we document the changes in the use of web

technologies by websites in the different cases outlined in Table 1 in Section 5.1. This lets us assess to what extent GDPR compliance is limited by territorial considerations of websites and users. Second, in Section 5.2, we switch perspectives to focus on web technology providers and the extent to which demand for their services changes following the GDPR. Third, we look at the changes in the structure of the web technology market to see whether some web technology vendors were able to improve their position following the GDPR (Section 5.3).

#### 5.1. Changes Within and Beyond the Territorial Scope of the GDPR

Any type of web technology that we can observe in our data can potentially be a tracking technology. When sending an HTTP request, websites are communicating with third-party servers. This communication can—as a technical principle—entail data, such as details of a user’s device (screen resolution, browser

Figure 2. Number of Requested Third-Party Domains



Notes. Average  $\log(y + 1)$  number of third-party domains to which websites with (non)EU audiences and (non)EU location send requests. As described in Section 2.1, the GDPR applies in all cases except non-EU/non-EU. Vertical line indicates the implementation of the GDPR on May 25, 2018. The vertical axis has a different scale in each panel. The purpose of this figure is to visualize the immediate changes (intercept) and changes in slopes after the GDPR for each case independently and not to compare differences in these changes across cases. Having a common scale would allow to compare the magnitudes of the changes after the GDPR across cases. This is what we do with the regression model in Equation (1) and Table 4.

version, operating system, etc.), but also personal data such as names, email addresses, and credit card details. Hence, when websites make less requests to third-party servers, this potentially enhances user privacy. Third-party requests that respond with cookies are by definition carrying personal data, because cookies are designed to identify the same user again by the same website or technology vendor at a different point in time. We therefore distinguish between requests to third-party domains overall and those to third-party domains that respond with at least one cookie. Although the former capture web technologies that could be and are used for collecting any kind of data, the latter reflect the de facto industry standard for collecting personal data online.

### 5.1.1. Web Technologies in General Are Used Less.

A plot of the raw data in Figure 2 shows that the average (log) number of third-party domains to which websites send requests follows an increasing trend interrupted at the time of the introduction of the GDPR. A reduction directly after the GDPR comes into force is clearly visible. However, the general trend seems largely unchanged. Thus, after the GDPR was enacted, websites did not decrease their third-party domain requests on a long-term basis. Rather, third-party requests are almost immediately back to their pre-GDPR trajectory.

Furthermore, Figure 2 suggests that the described changes happen in all cases—even in the case of non-EU audiences and a non-EU location, where the GDPR does not apply *de jure*. Looking at the post-GDPR trend, for websites outside of the EU (right column of Figure 2), the return to pre-GDPR levels is faster. At the end of our observation period in mid-November 2018, the number of requested third-party

domains is either back to the level just before the GDPR came into force (Case 3) or has increased far beyond that level (Case 4, Brussels effect). Visual inspection of the four panels also shows that the magnitude of the immediate change after the GDPR is largest in the standard case (Case 1), and smallest, yet still quite prominent in Case 4.

As noted previously, however, using the regression model specified in Equation (1), we put full focus on quantifying the immediate changes after the introduction of the GDPR and do neither interpret nor report estimates of trend changes. Using website fixed effects lets us control for time-invariant unobserved heterogeneity. Furthermore, because the model includes separate linear time trends for the pre- and post-GDPR periods, we can estimate the change in the intercept. That is, we can effectively compare the number of requested third-party domains just before and after the GDPR.

The estimates in column (1) of Table 4 suggest that in Case 1, we see a change of  $-7.6\%$  (90% confidence interval (CI):  $-8.0\%$ ,  $-7.2\%$ ) in the number of requested third-party domains immediately after the GDPR.<sup>17</sup> The estimate of the change in Case 2 is  $-4.9\%$  (90% CI:  $-5.5\%$ ,  $-4.3\%$ ). The estimates in column (2) of Table 4 imply a change in Case 3 of  $-4.5\%$  (90% CI:  $-5.0\%$ ,  $-4.1\%$ ) directly after the GDPR. Finally, our estimate of the change in Case 4 is  $-2.2\%$  (90% CI:  $-2.4\%$ ,  $-2.0\%$ ).

### 5.1.2. Web Technologies with Cookies Are Used Less.

Columns (3) and (4) of Table 4 repeat the same analysis but for the subset of third-party domains that send at least one cookie. The estimates in column (3) suggest the change in Case 1 is  $-10.4\%$  (90% CI:  $-10.9\%$ ,  $-9.8\%$ ). Regarding Case 2, we estimate  $-6.5\%$  (90% CI:

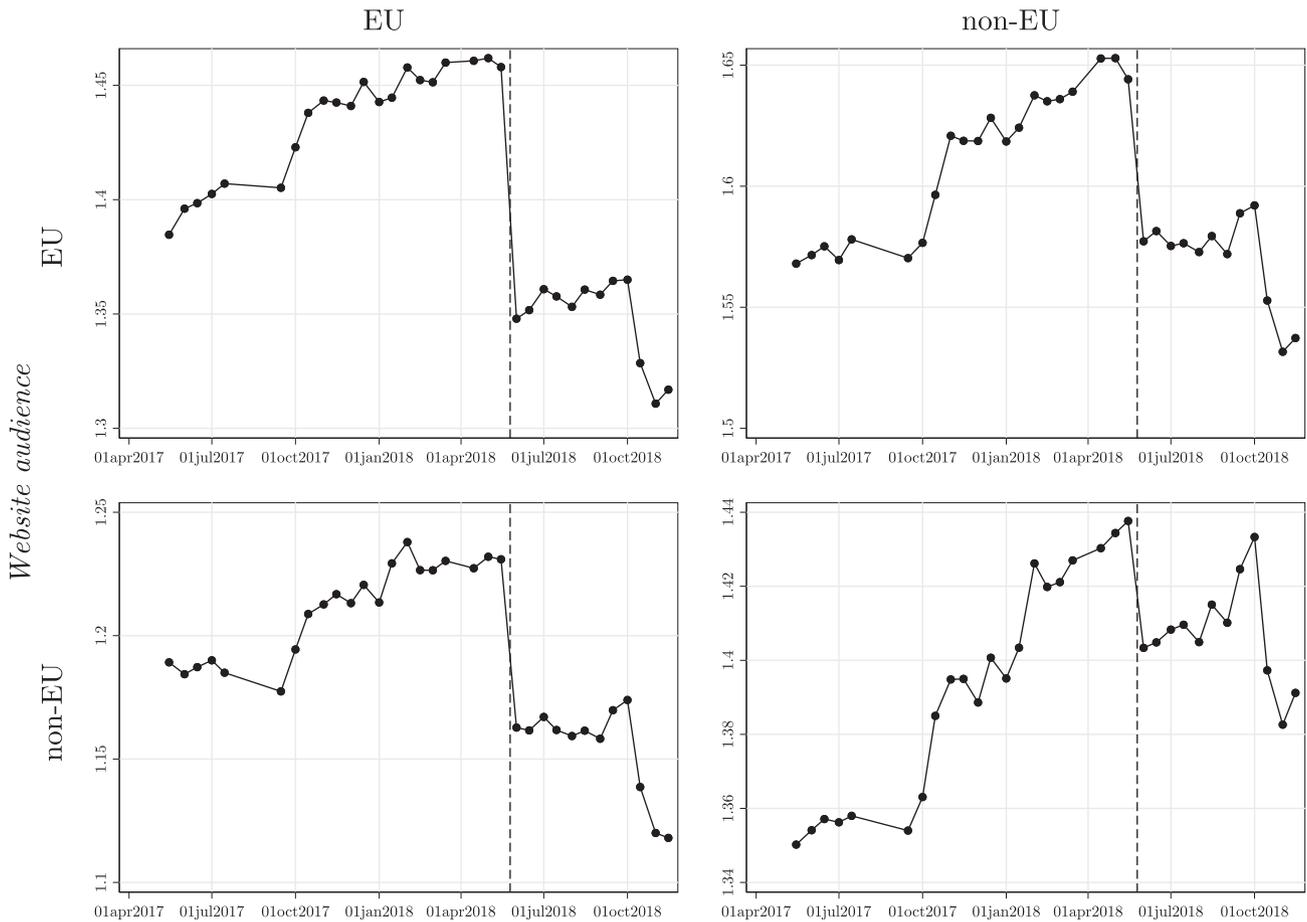
**Table 4.** Change in Number of Requested Third-Party Domains and Cookies

	Requests		Cookies	
	(1) EU firm	(2) Non-EU firm	(3) EU firm	(4) Non-EU firm
<i>Post</i> × <i>EU audience</i>	-0.0789*** (0.0026)	-0.0464*** (0.0028)	-0.1096*** (0.0036)	-0.0722*** (0.0042)
<i>Post</i> × <i>non-EU audience</i>	-0.0504*** (0.0038)	-0.0220*** (0.0011)	-0.0674*** (0.0051)	-0.0310*** (0.0015)
Observations	961,587	2,691,711	961,587	2,691,711
$\bar{R}^2$	0.8871	0.9054	0.8444	0.8641
Mean dependent variable	2.3706	2.4647	1.3476	1.4296

*Notes.* Dependent variable in columns (1) and (2) is the  $\log(y + 1)$  number of third-party domains that the website-host requests and in columns (3) and (4) the  $\log(y + 1)$  number of third-party domains that the website-host requests and respond with a cookie. *Post* indicates the period after May 25, 2018. *EU Audience* indicates whether a website host has a top-level domain that is specific to a country in the EU and/or whether the website's language is in a EU language and/or whether the website appears on any EU country's Alexa ranking. *Non-EU Audience* indicates the opposite. *EU firm* indicates that the website domain is associated with a firm that lists a headquarter in the EU on Crunchbase and/or a website where the Whois records indicate an owner with an address in the EU. Separate linear time trend for EU audience/non-EU audience before/after and website host fixed effects in all specifications. Standard errors in parentheses, clustered on the website-host-level.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

**Figure 3.** Number of Requested Third-Party Domains That Respond with Cookies



*Notes.* Average  $\log(y + 1)$  number of third-party domains to which websites with (non)EU audiences and (non)EU location send requests and that respond with cookies. The GDPR applies in all cases except non-EU/non-EU. Vertical line indicates the implementation of the GDPR on May 25, 2018. The vertical axis has a different scale in each panel. The purpose of this figure is to visualize the immediate changes (intercept) and changes in slopes after the GDPR for each case independently and not to compare differences in these changes across cases. Having a common scale would allow to compare the magnitudes of the changes after the GDPR across cases. This is what we do with the regression model in Equation (1) and Table 4.

[−7.3%, −5.7%]) less requested third-party domains with cookies directly after the implementation of the GDPR. Column (4) suggests that the change in Case 3 is −7.0% (90% CI: −7.6%, −6.3%) directly after the GDPR. We find that the change in Case 4 is −3.1% (90% CI: −3.3%, −2.8%).

The different magnitudes of the immediate changes after the GDPR across the different cases are also evident in Figure 3. However, it is striking that the increasing trend in the pre-GDPR period does not continue after the sudden and immediate drop. This is in sharp contrast to the overall number of requested third-party domains in Figure 2. This suggests that, after the GDPR went into force, a new state is reached in which less third-party cookies are used. The patterns are similar across the three cases in which the GDPR legally applies. However, the bottom right panel of Figure 3 implies that websites located outside the

EU that are catering a non-EU audience return to the levels directly before the GDPR, at least until October 2018.

We now discuss some alternative explanations for these findings. The apparent structural change after October 2018, with a further reduction in the number of requested domains that send at least one cookie, is visible in all four panes of Figure 3. Although it is difficult to pin down the underlying causal mechanism, one might speculate that it could be related to strategic behavior in response to (self-)regulatory pressure. On September 17, 2018, *Apple* released Safari 12, which included its *Intelligent Tracking Prevention* technology (ITP 2.0) that effectively bans third-party cross-site tracking cookies.<sup>18</sup> Similarly, *Firefox* released a new version of their browser on October 28, 2018, that blocked third-party cookies by default.<sup>19</sup> The way *HTTPArchive* collects data on third-party

requests and cookies does not rely on the technologies of Safari or Firefox; hence, the drop we observe cannot purely be a measurement issue. Rather, it suggests that there was a strategic reaction of websites and/or technology vendors.<sup>20</sup>

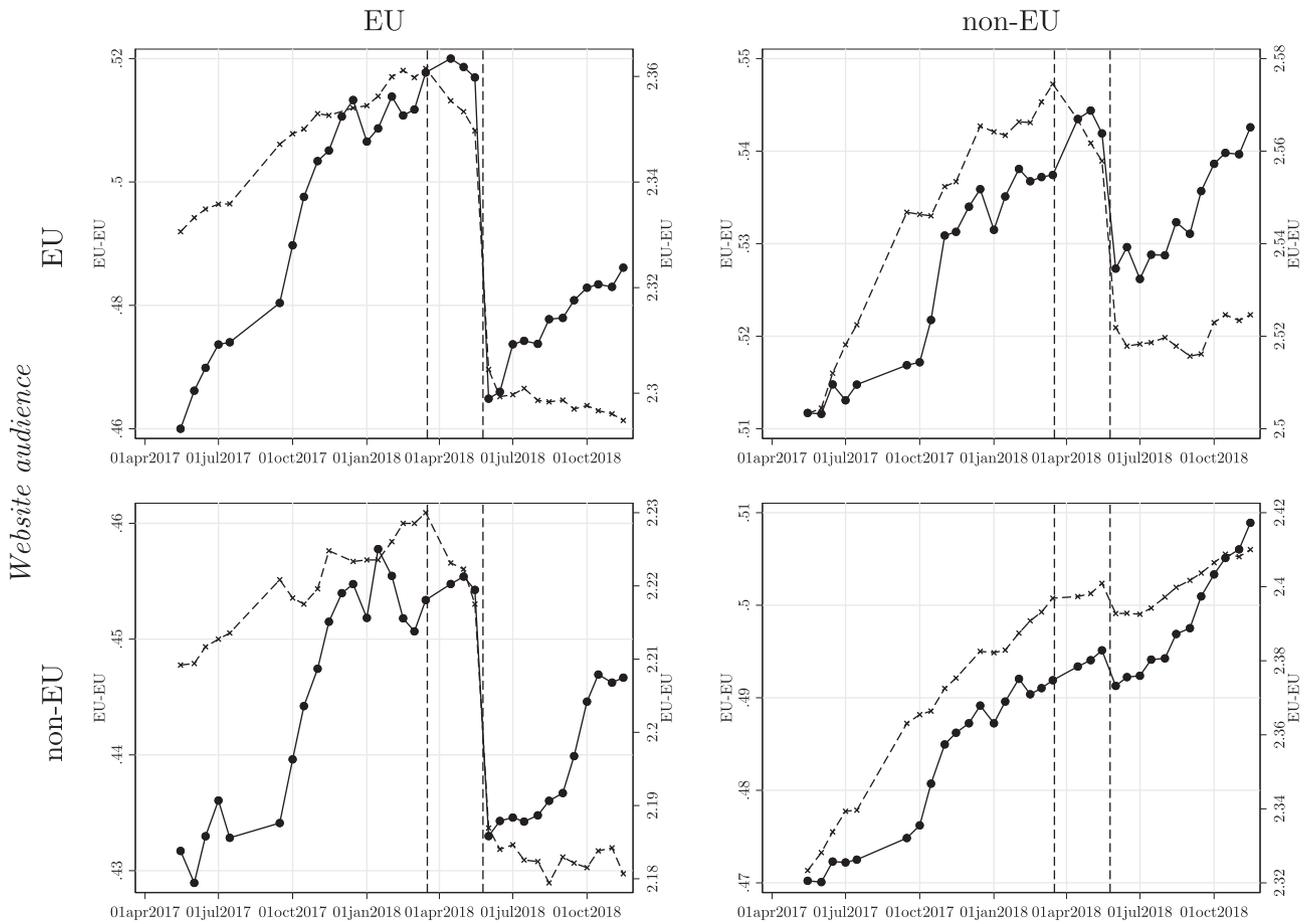
Although the GDPR was a major policy change with direct implications for the nearly 450 million inhabitants of the EU and beyond, there might be other events that coincide with the timing of its introduction. In particular, one might be worried the Facebook’s major privacy scandal (“Cambridge Analytica”) could affect our results. The public became aware of the Cambridge Analytica case with the revelations in a news article published on March 17, 2018. In Figure 4, we compare requests to Facebook-owned and non-Facebook third-party domains. Although requests to Facebook seem to decrease in all markets—except in Case 4 in Table 1—immediately after the Cambridge Analytica revelations, they decrease

much more heavily in all markets directly after the GDPR. Even if our data and research design does not allow us to make strong causal claims, we are therefore confident that our results concerning Facebook are not entirely driven by the Cambridge Analytica scandal.

**5.1.3. Mostly Less Popular Websites Change Operations After the GDPR.**

The results reported so far mask an important source of heterogeneity. As we show in Table 5, it is mostly less popular websites that change operations after the GDPR. Our measure of website popularity by its *global* Alexa rank, which is based on overall traffic.<sup>21</sup> We distinguish between the most popular websites (in the top 1000 most popular websites, which includes 599 websites in our sample) and lower ranked websites. The results in Table 5 indicate that the change after the GDPR that we document in the main text—both in terms of the number of requested

**Figure 4.** Requested Third-Party Domains, Facebook vs. non-Facebook



Notes. Average  $\log(y + 1)$  number of third-party domains to which websites with (non)EU audiences and (non)EU location send requests. The GDPR applies in all cases except non-EU/non-EU. The dashed plot indicates requests to Facebook’s servers; the solid plot indicates requests to servers of all other vendors. The first vertical line indicates the beginning of Facebook’s privacy scandal (Cambridge Analytica) with the revelations in a news article published on March 17, 2018. See <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. The second vertical line indicates the implementation of the GDPR on May 25, 2018. The vertical axis has a different scale in each panel. The purpose of this figure is to visualize the immediate changes (intercept) and changes in slopes after the GDPR for each case independently and not to compare differences in these changes across cases.

**Table 5.** Change in Number of Requested Third-Party Domains and Cookies: Website Popularity

	Requests		Cookies	
	(1) EU firm	(2) Non-EU firm	(3) EU firm	(4) Non-EU firm
<i>Post</i> × <i>EU Audience</i>	-0.0791*** (0.0026)	-0.0473*** (0.0029)	-0.1097*** (0.0036)	-0.0734*** (0.0042)
<i>Post</i> × <i>Non-EU Audience</i>	-0.0505*** (0.0038)	-0.0221*** (0.0011)	-0.0676*** (0.0051)	-0.0311*** (0.0015)
<i>Post</i> × <i>EU Audience</i> × <i>Top1k</i>	0.0312 (0.0306)	0.0363** (0.0160)	-0.0027 (0.0395)	0.0433* (0.0232)
<i>Post</i> × <i>Non-EU Audience</i> × <i>Top1k</i>	-0.0010 (0.0727)	0.0369 (0.0296)	0.0885 (0.0943)	0.0245 (0.0346)
Observations	960,993	2,689,302	960,993	2,689,302
$\bar{R}^2$	0.8870	0.9054	0.8444	0.8641
Mean dependent variable	2.3707	2.4651	1.3476	1.4300

*Notes.* Dependent variable in columns (1) and (2) is the  $\log(y + 1)$  number of third-party domains that the website host requests and in columns (3) and (4) the  $\log(y + 1)$  number of third-party domains that the website host requests and respond with a cookie. *Post* indicates the period after May 25, 2018. *EU Audience* indicates whether a website host has a top-level domain that is specific to a country in the EU and/or whether the website’s language is in a EU language and/or whether the website appears on any EU country’s Alexa ranking. *Non-EU Audience* indicates the opposite. *EU Firm* indicates that the website domain is associated with a firm that lists a headquarter in the EU on Crunchbase and/or a website where the Whois records indicate an owner with an address in the EU. *Top1k* indicates whether the website is ranked in the global top 1,000 most visited websites according to Alexa ( $n = 599$ ). Separate linear time trend for EU Audience/non-EU Audience before/after and website host fixed effects in all specifications. Standard errors in parentheses, clustered on the website-host-level.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

third-party domains as well as the number of third-party domains that respond with cookies—mostly comes from lower ranked websites. The triple interaction of postintroduction, audience type, and website popularity is either very close to zero, imprecisely estimated, or even significantly positive. In Section A.1.1 of the online appendix and in Table A.3, we show that a weighted regression leads to the same conclusion. This is consistent with the notion that small and new firms are disproportionately affected by privacy regulation (Campbell et al. 2015). We discuss the implications of this finding in more detail.

Similarly, as we discuss in more detail in Section A.1.1 of the online appendix, we find also find that general-audience websites reduce their requests to

third-party domains (and cookies) more than specific websites after the GDPR. This is in line with prior research on privacy regulation and effects on online advertising (Goldfarb and Tucker 2011).

## 5.2. Changes in the Collection of Personal Data

We now switch perspectives to focus on web technology vendors and investigate whether the web technology market has changed after the introduction of the GDPR. We are interested in whether the changes we have observed and quantified so far are especially pronounced among vendors that should be affected more strongly by GDPR: vendors disclosing that they collect and share personal data. We use detailed information on the data collection and sharing policies of

**Table 6.** Change in Websites Served by Data Type

	(1)	(2)	(3)
<i>Post</i>	0.013738 (0.020129)	0.013287 (0.020087)	0.015596 (0.019912)
<i>Post</i> × <i>Collect Personal</i>	-0.056821** (0.026621)	-0.062199** (0.029627)	-0.066829** (0.029267)
<i>Post</i> × <i>Share Personal</i>		0.014936 (0.036490)	-0.061624 (0.189001)
<i>Post</i> × <i>Collect Personal</i> × <i>Share Personal</i>			0.082506 (0.192486)
Observations	68,512	68,512	68,512
$\bar{R}^2$	0.951267	0.951313	0.951327
Mean dependent variable	2.712522	2.712522	2.712522

*Notes.* Dependent variable is the log the number of websites that send requests to third-party domains to firm  $i$ . *Post* indicates the period after May 25, 2018. Information on whether firms collect and/or share personal data comes from *Evidon*. All specifications include group-specific linear trends and firm-fixed effects. Standard errors clustered on the firm-level in parentheses.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

3,993 technology vendors via an industry database by Evidon. We first ask whether the number of requests websites send to a particular vendor changes at different rates after the GDPR depending on the vendor disclosing in their privacy policy that they collect or share personal data with third parties. To do this, we use information on privacy policies from September 2017, that is, before the GDPR. The results in column (1) of Table 6 show that firms that have indicated that they collect personal data prior to the GDPR receive about 6% less requests after the GDPR. This holds true when we additionally control for whether a firm discloses that they share personal data with third parties in column (2). There is no significant difference between firms that disclose sharing of personal data and those that do not. Finally, in column (3), we add the interaction between collection and sharing of personal data. This would capture firms disclosing that they do both. Again, there is no significant difference. Hence, we conclude that the reduction in requests we identified in our main results seems to come mostly from firms that disclose that they collect personal data.

An additional set of analyses on the dynamic changes in vendors' privacy policies is reported in the online appendix in Section A.1.2. Specifically, we find that vendors are more likely to disclose the collection and sharing of personal data, as well as a data retention policy after the GDPR went into effect. These results imply that the vendors change their approach to privacy rather than the previous results being driven by a change in the composition of firms.

### 5.3. Changes in the Competitive Landscape of Web Technologies

We now investigate changes in the market structure of web technology vendors. The results thus far suggest that websites reduce their interactions with third-party vendors after the introduction of the GDPR. Here we ask whether these changes are universal or whether some vendors are affected more or less strongly by websites' apparent reactions to the introduction of the GDPR. We do this because the related theoretical literature suggests that privacy protection can affect firms of different sizes differently, which can have implications for market structure (Campbell et al. 2015, Gopal et al. 2018). We start with an aggregate view, then look closer at firm-level measures of market power. In doing so, we will pay special attention to comparing the dominant firm to the rest of the market.

**5.3.1. Changes in Overall Market Structure.** We investigate whether the market structure in web technology markets has changed after the introduction of the GDPR. Using metadata from *whotracks.me*, we can

observe the ownership structure of third-party domains. For example, Google operates a variety of different domains, including less obvious domains such as *doubleclick.net*, *invitemedia.com*, and *2mdn.net*.

We measure market structure using the Herfindahl-Hirschmann index (HHI, the sum of squared market shares of all firms). We define the market share of a firm based on the sum of market shares for each third-party domain of that firm, that is,  $s_i = \sum_{j=1}^J (n_{i,j} / \sum_{k=1}^N n_{k,j})$ , where  $n_{i,j}$  is the number of websites that send requests to third-party domain  $j$  of web technology firm  $i$ ,  $N$  is the total number of third-party domains to which at least one website sends one request, and  $J$  is the number of third-party domains of firm  $i$ . Our definition throughout the rest of the paper does not make a distinction between websites according to their audience or firm location as defined in Table 1. In light of the results presented in Section 5.1 showing that firms change their interactions with third-party vendors globally, we believe that a market structure analysis should be carried out a global level rather than at the level of the legal applicability of the GDPR.<sup>22</sup>

With the caveat of a very small sample because of the high level of aggregation, we run a regression to investigate whether the HHI has changed after the introduction of the GDPR while controlling for an overall trend. In column (1) of Table 7, we find that HHI has increased by 16.0 points from a base of 950.8, which amounts to 1.68%. In column (2) of Table 7, we look at the hypothetical market structure without the firm with the largest market share (Google). Google's dominance becomes apparent when comparing the average pre-GDPR HHI when we include Google (950.8) relative to average pre-GDPR HHI when we exclude Google (70.1). In a hypothetical market without Google, the estimates are negative but imprecise.

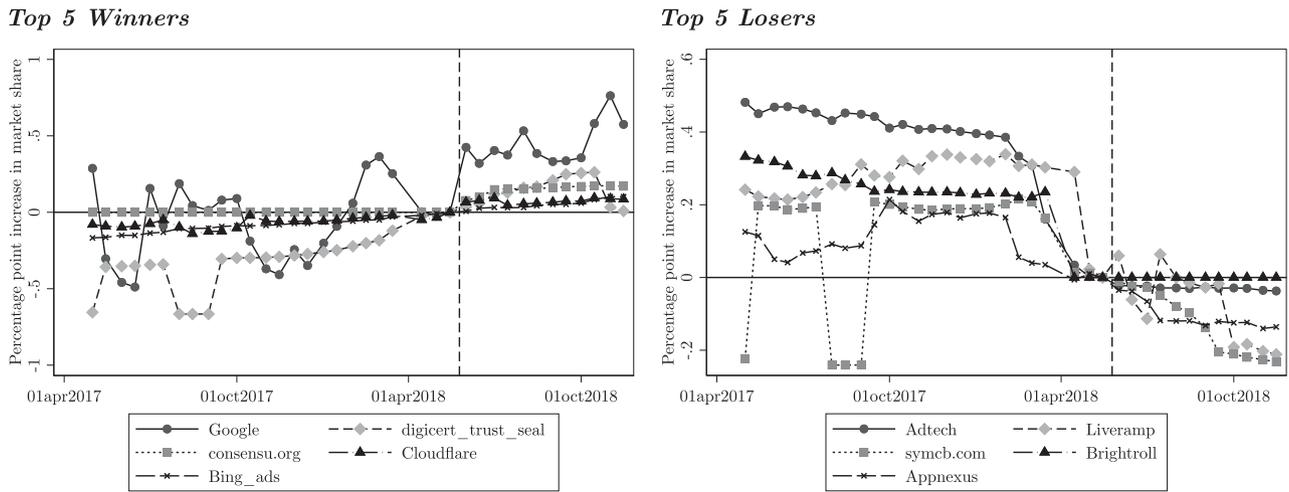
**Table 7.** Change in Market Structure of the Web Technology Industry

	(1) HHI all	(2) HHI without Google
After	16.016** (6.517)	-0.291 (0.606)
Observations	37	37
Pre-GDPR mean	950.833	70.090

*Notes.* Dependent variable is HHI, calculated as squared sum of market shares defined as  $s_i = \sum_{j=1}^J (n_{i,j} / \sum_{k=1}^N n_{k,j})$ , where  $n_{i,j}$  is the number of websites that send requests to third-party domain  $j$  of web technology firm  $i$ ,  $N$  is the total number of third-party domains to which at least one website sends one request, and  $J$  is the number of third-party domains of firm  $i$ . *Post* indicates the period after May 25, 2018. Separate linear time trend for before/after the GDPR. White-robust standard errors in parentheses.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

Figure 5. Winners and Losers



Notes. Top five firms with largest increase/decrease in market share at the end of our sample period compared with one snapshot before the implementation of GDPR on May 25, 2018. Market share is defined as  $s_i = \sum_{j=1}^J (n_{ij} / \sum_{k=1}^N n_{kj})$  where  $n_{ij}$  is the number of websites that send requests to third-party domain  $j$  of web technology firm  $i$ ,  $N$  is the total number of third-party domains to which at least one website sends one request, and  $J$  is the number of third-party domains of firm  $i$ .

**5.3.2. A Closer Look at Google.** Having established that Google is the dominant firm in the market and that changes in aggregate market concentration largely depend on Google, we now look at more granular data. We compare the changes in the market share of Google to changes in the market shares of all other firms. The definition of market share remains as described previously.

**5.3.2.1. Google Is the Biggest Winner in Terms of Market Share.** These results indicate that dominant firms in web technology markets play a special role and suggest that Google’s market shares have changed in a different way than the market shares of other firms. To provide a more detailed picture, we give a disaggregated descriptive view of market shares. Figure 5 plots the evolution of market shares of the five firms with the largest increases (“winners”) and of the five firms with the largest decreases (“losers”), relative to their market share level just before the introduction of the GDPR. The biggest winner is Google, and it is worth noting that among the winners there are three services that specialize in data security and consent for data collection, processing, and sharing (Consensu, Digicert Trust, and Cloudflare). The list of losers includes some of Google’s competitors in the advertising market (Adtech, Appnexus, Brightroll, Liveramp).

**5.3.2.2. Google Serves Less Websites Overall but Increases Its Market Share.** Although the previous illustration of the five largest firms is insightful, it can mask heterogeneity in the tails of the size distribution.

We therefore dig deeper and compare Google to all other vendors in a regression setup.

First, and to provide further detail on how Google’s market share increased after the introduction of the GDPR, we look at the numerator in the definition of market share. We run a regression comparing Google to all other vendors. We look at the number of websites served by a web technology vendor as well as its market share while controlling for group-specific time

Table 8. Change in Levels and Market Shares: Google vs. All Other Firms

	(1) No. websites	(2) Market shares
<i>Post</i> × <i>Google</i>	−0.0261*** (0.0000)	0.2731*** (0.0000)
<i>Post</i> × <i>Non-Google</i>	0.0006 (0.0013)	−0.0000 (0.0000)
Observations	2,341,804	2,341,804
$\bar{R}^2$	0.8847	0.9994
Pre-GDPR Google	13.1370	29.6763
Pre-GDPR non-Google	0.6155	0.0011

Notes. Dependent variable in column (1) is the  $\log(y + 1)$  number of websites served by all services of web technology firm  $i$  combined, in column (2) is the market share defined as  $s_i = \sum_{j=1}^J (n_{ij} / \sum_{k=1}^N n_{kj})$ , where  $n_{ij}$  is the number of websites that send requests to third-party domain  $j$  of web technology firm  $i$ ,  $N$  is the total number of third-party domains to which at least one website sends one request, and  $J$  is the number of third-party domains of firm  $i$ . *Post* indicates the period after May 25, 2018. *Google* indicates the joint levels/market share of all domains of Google. *Non-Google* indicates the average levels/market share of all other firms. All specifications include firm-fixed effects and separate linear time trends. Standard errors in parentheses clustered on the firm-level.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

trends and firm-specific unobserved heterogeneity. Column (1) of Table 8 shows that the total number of websites that send requests to all Google services combined decreases by about 2.4% after the introduction of the GDPR. This refers to a pre-GDPR log number of websites of 13.14 (or 508,896 websites, equivalent to the average website in our sample sending about 4.6 requests to Google services). For the average non-Google firm, we do not find a significant change after the introduction of the GDPR. Looking at the pre-GDPR mean for non-Google vendors (0.61 in logs, i.e., 1.8 websites), Google's strong market position becomes evident. Looking at market shares in column (2) of Table 8, we find a significant increase of Google services, suggesting that Google's market share across all its services increased by about 0.3 percentage points from a base of 29.7% before GDPR. The market share of the average non-Google firm (again across all the firm's services) does not change significantly after GDPR. Again, Google's strong market position becomes clear when looking at the pre-GDPR market share of 0.001% of the average non-Google firm. A likely explanation for the results in Table 8 is that, although most firms received less third-party requests after the introduction of the GDPR became effective, some firms—and most strikingly Google—lose relatively less such that their market shares increase after the GDPR.

**5.2.2.3. Changes Are Most Pronounced in Google's Key Markets.** Google is one of few web technology vendors in our sample that operate in multiple submarkets, that is, offer services across different categories. Some of these services may be less or more strongly affected by the introduction of the GDPR, for example, because they (do not) involve personal data.

To further explore the special role of Google after the GDPR, we therefore differentiate between submarkets in which Google is active. Using data from *whotracks.me*, we classify third-party domains in the five categories of *Video/Audio Players*, *Advertising*, *Analytics*, *CDN/API*, and *Other/Unknown* services. We accordingly change our definition of market share to the category-level. Hence,  $s_{i,c} = \sum_{j=1}^{J_c} (n_{i,c,j} / \sum_{k=1}^{N_c} n_{k,c,j})$ , where  $n_{i,c,j}$  is the number of websites that send requests to third-party domain  $j$  in category  $c$  of web technology firm  $i$ ,  $N_c$  is the total number of third-party domains in category  $c$  to which at least one website sends one request, and  $J_c$  is the number of third-party domains in category  $c$  of firm  $i$ .

In Table 9, we look at changes in the  $\log(y + 1)$  number of websites that send requests to Google-owned and non-Google services, distinguishing between different service categories. Again, note the large difference in pre-GDPR means. Google's web technologies are, on average, requested by thousands of websites, whereas average non-Google web technologies are

requested by less than 10 websites. Google's largest market is *Analytics*: the pre-GDPR mean is about 212,011, indicating that many of the 110,706 websites in our sample send more than one request to Google's domains in that category. Put differently, the websites in our sample send about 1.9 requests on average to Google's analytics services. Looking at the changes after the GDPR, we see that the number of websites requesting Google services decreases in all categories except *Other/Unknown* by between 1.3% and 4.1%. We also see significant decreases in *Advertising*, *Analytics*, and *CDN/API* for non-Google services, ranging between 2.7% and 4.9%. The point estimates of Google's decrease in the number of websites served is smaller than the decrease of non-Google services in the *Advertising*, *Analytics* and *Other/Unknown* categories, but much larger in *Video/Audio Players* and *CDN/API*. These results suggest that, although the pie is shrinking for everybody, some firms walk away with relatively large pieces.

Finally, we look at market share changes in Table 10. Again, we distinguish between Google and non-Google services and look separately at the same five categories. We only see significant changes in the market shares of Google, and not of other firms. In *Video/Audio Players*, Google's market share decreases by about 1.1 percentage points from a pre-GDPR mean of about 75.8% and in *CDN/API* by about 0.4 percentage points from mean of about 70.6%. However, in *Advertising*, Google's market share increases by about 0.4% from pre-GDPR base of about 27.1%. In the *Analytics* category, we observe an increase in Google's market share of 1.4 percentage points from a base of 39%, and our results suggest that Google's market share increased by about 0.03 percentage points in the *Other/Unknown* category from a base of 1.2%. Hence, Google's market position has improved after the GDPR, considering that the overall market has shrunk. This is especially the case in the categories of *Advertising* and *Analytics*, technologies that collect, process, and sometimes share personal data.

The analyses in Tables 9 and 10 may be driven by a large tail of third-party domains that receive only a small number of requests from the websites in our sample. In the online appendix (Tables A.11 and A.12), we show that the results are robust to removing all third-party domains that receive less than 50 requests in a given time period.

#### 5.4. Limitations and Robustness

Given the institutional setting of the GDPR, it is difficult to establish causal estimates. We discuss this limitation here. Conditional on the issue of causality, however, our results are robust to alternative variable definitions and a more flexible approach to modeling the dynamics of

**Table 9.** Change in Levels by Submarket

	(1) Video	(2) Advertising	(3) Analytics	(4) CDN/API	(5) Other
<i>Post</i> × <i>Google</i>	-0.012557*** (6.1e-18)	-0.040471*** (1.4e-17)	-0.01697*** (6.1e-17)	-0.023483*** (4.1e-18)	0.013068*** (4.6e-17)
<i>Post</i> × <i>Non-Google</i>	-0.002325 (0.002143)	-0.047413*** (0.014049)	-0.026433*** (0.005668)	-0.002923** (0.001224)	-0.006726 (0.005318)
Observations	33,344	33,344	33,344	33,344	33,344
$\bar{R}^2$	0.986741	0.975713	0.983701	0.997838	0.982263
Pre-GDPR Google	10.1852	12.2644	11.4786	11.7228	6.69627
Pre-GDPR non-Google	0.1018	2.25412	0.921685	0.186973	0.40283

*Notes.* Dependent variable is the log the number of websites that send requests to services owned by web technology firm  $i$ . *Post* indicates the period after May 25, 2018. *Google* indicates the joint levels/market share of all domains of Google. *Non-Google* indicates the average levels/market share of all other firms. All specifications include firm-fixed effects and separate linear time trends. Standard errors clustered on the firm-level in parentheses.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

third-party requests. For details, see Section A.2 of the online appendix.

**5.4.1. Definitions and Measures.** Our point estimates of the changes after the introduction of the GDPR appear conservative compared with a large number of alternative ways of measuring the target audience of websites and/or the geographical location of the organization that operates them. We describe these alternative definitions and measures in Section A.2.1 of the online appendix. Our baseline results are either statistically not distinguishable or more conservative than the results obtained in 44 other specifications. Most importantly, in all 96 models we estimate, we always find a significant decrease in the number of requested third-party domains or the number of requested third-party domains that respond with cookies. Our results also hold for websites targeting multiple international audiences.

**5.4.2. Alternative Data Source for Third-Party Requests.** A potential limitation of *HTTPArchive* is that the servers that crawl websites to collect their

interactions with third-party domains are physically located in California. If websites used different tracking technology depending on user location, we might mismeasure the changes after the introduction of the GDPR. In Section A.2.2 of the online appendix, we show that our results hold when using a different data source that does not rely on one specific user location. Using data on third-party requests collected through a browser plugin (*whotracks.me*), and therefore recording the third-party requests of a website as the user navigates to it, we can differentiate between users in the US and users in the EU (Table A.10). This data set has its own shortcomings, but we arrive at the same baseline conclusion: Websites visited by users both inside and outside the EU reduce the number of interactions with third-party web technology providers after the introduction of the GDPR.

**5.4.3. Alternative Functional Form for Modeling Dynamics.** The specification in Equation (1) assumes that the dynamics of third-party requests can be linearly approximated. As we show in the online

**Table 10.** Change in Market Shares by Submarket

	(1) Video	(2) Advertising	(3) Analytics	(4) CDN/API	(5) Other
<i>Post</i> × <i>Google</i>	-1.09083*** (6.5e-16)	0.392191*** (1.1e-17)	1.44175*** (1.3e-17)	-0.38892*** (2.2e-16)	0.025299*** (6.8e-18)
<i>Post</i> × <i>Non-Google</i>	0.001048 (0.000873)	-0.000377 (0.001053)	-0.001385 (0.001352)	0.000374 (0.000767)	-0.000024 (0.002691)
Observations	33,344	33,344	33,344	33,344	33,344
$\bar{R}^2$	0.999544	0.998629	0.997921	0.999829	0.985293
Pre-GDPR Google	75.7562	27.0608	39.4429	70.6164	1.19692
Pre-GDPR non-Google	0.023289	0.070066	0.058172	0.028226	0.094912

*Notes.* Dependent variable is market share defined as  $s_{i,c} = \sum_{j=1}^{J_c} (n_{i,c,j} / \sum_{k=1}^{N_c} n_{k,c,j})$ , where  $n_{i,c,j}$  is the number of websites that send requests to third-party domain  $j$  in category  $c$  of web technology firm  $i$ ,  $N_c$  is the total number of third-party domains in category  $c$  to which at least one website sends one request, and  $J_c$  is the number of third-party domains in category  $c$  of firm  $i$ . *Post* indicates the period after May 25, 2018. *Google* indicates the joint levels/market share of all domains of Google. *Non-Google* indicates the average levels/market share of all other firms. All specifications include firm fixed effects and separate linear time trends. Standard errors clustered on the firm-level in parentheses.

\* $p < 0.10$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

appendix in Section A.2.3, we obtain similar results when we allow for a more flexible functional form for the dynamics. In particular, comparing just the last observation before period (May 1, 2018) to the first observation of the after period (June 1, 2018) in what resembles a regression-discontinuity-in-time design, we obtain qualitatively similar estimates of the immediate change after the GDPR as in the baseline specification reported in Table 4.

#### 5.4.4. Potential Confounding Events and Anticipation.

Although the GDPR was a major policy change with direct implications for the nearly 450 million inhabitants of the EU and beyond, we are careful about making causal claims in our empirical analysis as other events may have coincided with the timing of the GDPR coming into force. As we show in the online appendix in Section A.2.4, we do not see significant changes in the interactions with third-party web technology providers when we focus on online piracy websites that operate in a legal gray area and should therefore not be bound by the introduction of the GDPR. Furthermore, we show that websites had not already initiated significant changes in their interactions with third-party web technology providers when the Council of the European Union and the European Parliament adopted the GDPR in April 2016.

## 6. Discussion

We now discuss how our findings from Section 5 relate to the institutional features of the GDPR laid out in Section 2.

### 6.1. Privacy Law and Dynamic Compliance Risks

We find that websites reduce their connections to web technology providers after the GDPR went into force, and that this reduction was particularly pronounced regarding requests involving cookies, which typically process personal data (see Section 5.1). These findings can be explained by the increasingly complex compliance landscape websites have to navigate after the GDPR. Given the risks created by the GDPR's joint responsibility regime (see Section 2.2), if websites cannot accurately assess the privacy risks originating from their web technology provider, the best way to reduce liability exposure is to reduce their use of web technology providers. We see that third-party web technology providers collecting personal data receive less requests after the enactment of the GDPR (see Section 5.2). We also find a persistent reduction in third-party domain requests involving cookies (see Section 5.1.2). Given that third-party cookies typically include personal data at least by identifying a browser and a device, this outcome aligns with one of the primary goals of the European legislator when creating

the GDPR: to increase the protection of individuals' privacy while maintaining the benefits of data processing, for example, by implementing a data minimization principle (Art. 5(1)(c), 25(1) and Recitals 78, 156 GDPR).

### 6.2. Privacy Law and Regulatory Competition

As described in Section 5.1, we observe that, at least temporarily, even websites located outside the EU and catering to a non-EU audience reduce their use of third-party web technology providers, especially cookies, after the GDPR. At first sight, it is counterintuitive for websites not legally bound by the GDPR to conform to its rules (Table 1). However, considering the broad territorial application of the GDPR (see Section 2.1), this de facto compliance is understandable for three reasons.

First, under general principles of international public law, the EU cannot regulate the processing of personal data that takes place outside of and is not related to the EU. Yet, the EU has expanded the de facto territorial reach of European privacy laws well beyond the geographical boundaries of the EU. As complying with the GDPR is costly, some global technology companies simply apply the GDPR to all their consumers worldwide, even though the GDPR does not require them to do so. These companies save costs by not having to offer two versions of their products and services (for users inside and outside the EU).<sup>23</sup> A second rationale for global adoption of the GDPR by websites can be that the GDPR has served as a role model for other countries to follow suit (Bradford 2020, p. 143–144). By adopting GDPR-compliant rules for all their consumers worldwide, firms may preempt the anticipated adoption of GDPR-like regimes in countries outside the EU. Third, non-EU websites that consider entering the EU market may decide to adopt EU rules right away: To realize the option value of being able to serve EU customers at some point in the future, a website may decide to comply with GDPR ex ante, rather than incurring the cost of creating and maintaining multiple versions of the website, one GDPR-compliant version for EU customers and one for the rest of the world (Frankenreiter 2022).

The literature on international regulatory competition has developed a general theory that the EU has de facto expanded some of its strict regulatory laws beyond its borders through a combination of market mechanisms and unilateral regulatory globalization. European privacy law is one example of the Brussels effect, according to Bradford (2012, 2020), in addition to European antitrust, consumer health, and environmental law (Goldsmith and Wu 2006, Schwartz 2019). Our study provides the first large-scale empirical evidence of this argument in European privacy law.<sup>24</sup>

The EU is not the only jurisdiction whose privacy laws may have extraterritorial reach (Rustad and Koenig 2019, Bygrave 2021). The California Consumer Privacy Act (CCPA)—which became effective January 1, 2020, and incorporates several GDPR concepts such as the rights of access, portability, and data deletion—applies to all for-profit entities doing considerable business in California. It protects California residents that are either in California for other than a temporary or transitory purpose or are domiciled in California but are currently outside the state for a temporary or transitory purpose (Cal. Civ. Code §1798.140(c), (g); Cal. Code Regs. tit. 18, §17014). Consequently, the CCPA can apply to websites and web technology providers located outside California or even outside the United States, as long as they cater to California residents. Several other U.S. states are considering introducing similar or even stricter privacy laws than the CCPA. As with the Brussels effect in the context of the GDPR, we may observe a *California effect* (Vogel 1995) in future U.S. privacy law, and state privacy laws may converge toward the jurisdiction with the strictest regulatory standard.

In fact, this debate transcends the boundaries of privacy regulation. In the spring of 2021, the European Commission proposed an “Artificial Intelligence Act” (AI Act), coordinating European rules on the legal and ethical implications of artificial intelligence (European Commission 2021). The AI Act proposes fines that go even beyond the sanctioning regime of the GDPR (up to 6% of the worldwide annual turnover, see Art. 71 (3) of the Draft AI Act), and it would apply if an AI system or its output are used in the European Union, regardless of whether the systems provider is located within the EU or not. The proposed AI Act encompasses important building blocks—in particular, an effective sanctioning regime and extraterritorial application—that could contribute to a Brussels effect in the context of AI regulation. In this context, our study sheds an empirical light on the mechanisms that may enable regulatory competition and a race-to-the-top in regulating privacy, AI, and other areas of digital society.

### 6.3. Privacy Law and Antitrust Policy

We find that, although the market for third-party web technologies shrank in the period following the GDPR’s enactment, Google’s position in various web technology markets improved relative to competitors (see Sections 5.3.1 and 5.3.2). Increasing concentration and an increasing market share of the dominant firm is most likely not what European legislators had in mind when designing the GDPR. Indeed, the European Commission stressed in 2012 how the procompetitive effects of the future GDPR would increase the attractiveness of Europe as a location to do business (European Commission 2012, p. 148–149). However, our findings are consistent with

the dynamics of data-intensive markets. In data-intensive markets, large firms may have an advantage in the processing of personal data. As mentioned in Section 2.3, the revised e-Privacy Directive and the GDPR require firms to gather user consent for using cookies and processing personal data. As long as the data stays within the firm, the firm may control its compliance risks by a firm-wide consent management system. Once data are shared with a third party, however, the firm must inform its consumers and may be jointly liable for privacy violations (see Section 2.2). Hence, the GDPR has created an environment in which data sharing within firm boundaries is less risky than data sharing across boundaries.

Moreover, in line with the compliance risks outlined in the preceding section, websites may choose large web technology providers over small ones because these may have more resources to weather legal challenges created by the GDPR. By choosing a large web technology provider, a website may therefore reduce its own compliance risk. The GDPR widely implemented and enforced the consent requirement for websites, which disproportionately benefits larger firms offering a broader range of services: as most Internet users have used some of Google’s services, Google could gather consent from most Internet users. Such consent then typically covers a broad range of Google products and services. The larger a service provider becomes, the cheaper it may become to gather broad user consent. This is consistent with theoretical work showing that larger firms can benefit disproportionately from access to user data (Farboodi et al. 2019, Hagiú and Wright 2020). Our results also speak to work showing that consent-based privacy regulation can disproportionately benefit firms offering a larger scope of services and that privacy regulation can increase market concentration by restricting data flows across firms (Campbell et al. 2015, Acemoglu et al. 2019, Jin and Wagman 2020, Jones and Tonetti 2020).

Although our results are in line with literature pointing out tradeoffs between privacy protection and competition, we can only provide empirical evidence regarding one side of the tradeoff: we document empirically the changes in third-party domain requests in general and cookies in particular after the GDPR has been enacted. However, we cannot quantify the costs of more concentration with our data.

## 7. Conclusion

We provide robust large-scale evidence on the changes occurring around the time when the GDPR came into force in the context of websites and web technology providers. We show how websites—with in a time frame of six months—reduce their compliance risks after the GDPR: they reduce the number of

third-party web technology providers they use, in particular relating to third-party cookies. We offer empirical evidence of the Brussels effect in European privacy law: Websites and web technology providers that are located outside the EU, cater to non-EU audiences, and are therefore not subject to the GDPR still comply with it. Finally, we demonstrate that, although markets for web technologies shrunk in size after the enactment of the GDPR, the dominant firm—Google—increased its market share vis-à-vis competing web technology providers. Our findings suggest that some of the key implications of the GDPR may not relate to privacy, but to antitrust policy and regulatory competition. Although such regulatory spillovers have general implications for debates on how to govern data and AI, we leave the implications for the theoretical relationship between privacy and antitrust laws to future research.

### Acknowledgments

Earlier versions of this paper circulated under the title “European Privacy Law and Global Markets for Data.” The authors thank the editor, the associate editor, and two reviewers for valuable comments and suggestions. The authors are grateful for the feedback of seminar and conference participants at Academy of Management Annual Meeting, Boston University Law, Center for Research in Economics and Management, Digital Economy Workshop Tel Aviv, Economics of Digitization Seminar Paris, ETH Zurich, European Association for Law and Economics Conference, IESE Business School, Louvain Economics of Digitization Online Seminar, Ludwig-Maximilians-Universität Munich, Minnesota Carlson Management Information Systems Seminar, National Bureau of Economic Research (NBER) Digitization Summer Institute, Paris Conference on Digital Economics, Platform Strategy Research Symposium, Rennes School of Business, Rotterdam School of Management, Tilburg University, University of Illinois, University of Zurich, Working Group Competition & Regulation Würzburg, Workshop on Information Systems and Economics (WISE), and Center for European Economic Research (ZEW) Conference on the Economics of Information and Communication Technologies. The authors thank Ilia Azizi, Daniel Bötticher, David Campi, and Ines de Planta for helpful research assistance.

### Endnotes

<sup>1</sup> When we refer to the *effects doctrine* and the *Brussels effect*, we refer to concepts as introduced in the legal discourse, not to *effects* in the sense of causal inference.

<sup>2</sup> For a recent U.K. High Court of Justice decision on these questions, see *Walter Tzvi Soriano v. Forensic News*, [2021] EWHC 56 (QB).

<sup>3</sup> In recent case law, the European Court of Justice held that websites and web technology providers can be responsible as “joint controllers” (Art. 26 GDPR) even if the website cannot control what personal data are transmitted to or processed by the web technology provider. See European Court of Justice, June 5, 2018, case C-210/16, *ULD Schleswig Holstein v. Wirtschaftsakademie Schleswig-Holstein* (Fan Page); July 29, 2019, case C-40/17, *Fashion*

*ID v. Verbraucherzentrale NRW* (Fashion ID); European Data Protection Board, 2021. In *Fashion ID*, the European Court of Justice established a joint controller responsibility between a website that used a Facebook Like button and Facebook. The Court stressed that the joint responsibility is limited to those steps of the data processing that were jointly determined.

<sup>4</sup> Such liability does not exist if the website can prove that it was not responsible for the violation. However, the website may often share some responsibility, or it may have difficulties proving otherwise in court (Articles 26(1), (3), 82(3), (4) GDPR). Also, the joint liability does not extend to administrative fines.

<sup>5</sup> The European Data Protection Board, for example, issued its guidelines on the territorial applicability of the GDPR six months after the GDPR had become effective (European Data Protection Board 2019a).

<sup>6</sup> According to industry reports, compliance costs for large U.K. firms (FTSE 350) were \$1.1 billion, and \$7.8 billion for large U.S. firms (Fortune 500); see <https://www.forbes.com/sites/oliversmith/2018/05/02/the-gdpr-racket-whos-making-money-from-this-9bn-business-shakedown>.

<sup>7</sup> Cookies are small files stored in the user’s web browser placed by a website when visited. They contain information about the user’s visit to provide additional functionality. Cookies placed by the visited websites are called first-party cookies, whereas cookies placed by other entities are called third-party cookies. These are typically used to track user behavior across different websites.

<sup>8</sup> In some cases, the GDPR also allows for processing without user consent. In particular, if the website has a prevailing legitimate interest to process personal data—for example, IT security measures or fraud prevention—Article 6(1)(f) GDPR allows the website to process such data without explicit consent of the user (see also Recitals 47 and 49 GDPR). For the types of technologies of interest for our study, user consent is usually required.

<sup>9</sup> Art. 5(3) e-Privacy Directive concerns the *storing* of a cookie on a user’s device, whereas Art. 6(1)(a) GDPR concerns the *processing* of personal data that has been gathered through cookies or other means. On the complex relationship between the GDPR and the e-Privacy Directive, see European Data Protection Board (2019b).

<sup>10</sup> See the description at <https://www.thirdrocktechkno.com/blog/how-to-choose-a-technology-stack-for-web-application-development/>.

<sup>11</sup> Detailed information for replication, including instructions on how to obtain the publicly available data we use is available at <https://github.com/cpeukert/gdpr>.

<sup>12</sup> Regarding our main estimation period, the public data set, available in Google’s BigQuery web service, does not contain information for 12-01-2016, 01-01-2017, 01-15-2017, and 04-01-2018.

<sup>13</sup> For our empirical analysis, we call *subdomain.domain.com* a *website-host* and *domain.com* a *website*.

<sup>14</sup> HTTPArchive has introduced several changes in the technology they use to collect information on HTTP requests and the list of website-hosts from which they collect information. This reduces the number of websites that we can consistently observe.

<sup>15</sup> Websites can send multiple requests to different URLs of the same third-party domain. For example, we may observe two requests to *google-analytics.com* for the same host. One request concerns the URL <https://www.google-analytics.com/r/collect> (with a number of parameters), and the second request concerns the URL <https://www.google-analytics.com/analytics.js>.

<sup>16</sup> Although the GDPR has affected the amount of information publicly listed in the domain registry, the country of the registrant remains available. See <https://www.icann.org/resources/pages/gtld-registration-data-specs-en>.

<sup>17</sup> Here and in the rest of the study, we approximate effect sizes as  $(\exp(\text{coeff}) - 1) \times 100$ . Although this is not accurate for the models where the dependent variable is  $\log(y + 1)$  transformed (Bellego et al. 2021), we show in the online appendix in Table A.1 that the results are robust to using a  $\log(y)$  transformation.

<sup>18</sup> See <https://www.macrumors.com/2018/09/17/apple-releases-safari-12-macos-high-sierra/> and <https://webkit.org/blog/8311/intelligent-tracking-prevention-2-0/> for a technical description of ITP 2.0.

<sup>19</sup> See <https://blog.mozilla.org/en/products/firefox/latest-firefox-rolls-out-enhanced-tracking-protection/>.

<sup>20</sup> For example, shortly after the announcement of Apple's first version of the ITP technology in June 2017, Google (in September 2017), Microsoft (in January 2018), and Facebook (in October 2018) adjusted their tracking technology such that they can "hide" behind the domain of the first-party website when issuing cookies. See <https://webkit.org/blog/7675/intelligent-tracking-prevention/>, <https://searchengineland.com/google-analytics-adwords-response-apple-intelligent-tracking-prevention-282233>, <https://about.ads.microsoft.com/en-us/blog/post/january-2018/conversion-tracking-update-on-bing-ads>, and <https://www.adweek.com/programmatic/the-facebook-pixel-will-add-a-first-party-cookie-option/>.

<sup>21</sup> Information on a website's global Alexa rank is included in the data set that we obtain from *HTTPArchive*. This measure is time-invariant and was obtained in October 2017, see <https://discuss.httparchive.org/t/alexa-rank-attribute-in-the-pages-table/1391>.

<sup>22</sup> In Figure A.6 in the online appendix, we additionally break out the HHI definitions by distinguishing between websites according to geographical audience/firm location defined in Table 1.

<sup>23</sup> In May 2018, Microsoft announced, for example, that it would apply the GDPR to all consumers worldwide; see <https://blogs.microsoft.com/on-the-issues/2018/05/21/microsofts-commitment-to-gdpr-privacy-and-putting-customers-in-control-of-their-own-data>. In November 2020, a senior Apple executive noted that the GDPR provides the privacy model for Apple around the world; see <https://www.euronews.com/2020/11/17/europe-s-gdpr-provides-our-privacy-model-around-the-world-apple-vp-says>; see also Bradford (2020, p. 143–144) and Rustad and Koenig (2019, p. 391–96).

<sup>24</sup> In an empirical study of privacy policies, Frankenreiter (2022) does not find signs of this argument. As that study points out on p. 57, the different findings may result from the fact that it is easier and less costly to create territory-specific privacy policies than territory-specific websites.

## References

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2019) Too much data: Prices and inefficiencies in data markets. Working paper, No. w26296 National Bureau of Economic Research, Cambridge, MA.

Ang S, Straub DW (1998) Production and transaction economies and is outsourcing: A study of the us banking industry. *Management Inform. Systems Quart.* 22(4):535–552.

Aridor G, Che Y-K, Salz T (2020) The economic consequences of data privacy regulation: Empirical evidence from GDPR. Preprint, submitted January 29, <https://dx.doi.org/10.2139/ssrn.3522845>.

Bellego C, Benatia D, Pape L-D (2021) Dealing with logs and zeros in regression models. *Série des Documents de Travail* No. 2019-13, Center for Research in Economics and Statistics (CREST), France.

Bessen J, Impik SM, Reichensperger L, Seamans R (2020) GDPR and the importance of data to AI startups. Preprint, submitted April 15, <https://dx.doi.org/10.2139/ssrn.3576714>.

Bradford A (2012) The Brussels effect. *Northwestern University Law Rev.* 107:1–68.

Bradford A (2020) *The Brussels Effect* (Oxford University Press, Oxford, UK).

Bygrave LA (2021) The 'Strasbourg effect' in data protection: Its logic, mechanics and prospects in light of the 'Brussels effect'. *Comput. Law Security Rev.* 40:105460.

Campbell J, Goldfarb A, Tucker C (2015) Privacy regulation and market structure. *J. Econom. Management Strategy* 24:47–73.

Cheng HK, Tang QC, Zhao JL (2006) Web services and service-oriented application provisioning: an analytical study of application service strategies. *IEEE Trans. Engrg. Management* 53(4):520–533.

Currie W (2000) The supply-side of it outsourcing: the trend towards mergers, acquisitions and joint ventures. *Internat. J. Phys. Distribution Logist. Management.* 30(3/4):238–254.

Dabrowski A, Merzdovnik G, Ullrich J, Sendera G, Weippl E (2019) Measuring cookies and web privacy in a post-GDPR world. *Proc. Internat. Conf. on Passive and Active Network Measurement* (Springer, Berlin), 258–270.

Dahl D, Meyer A, Neely M (2016) *Bank Size. Compliance Costs and Compliance Performance in Community Banking* (Federal Reserve Bank of St. Louis, St. Louis, MO).

Degeling M, Utz C, Lentzsch C, Hosseini H, Schaub F, Holz T (2019) we value your privacy ... now take some cookies: Measuring the GDPR's impact on web privacy. *Informatik Spektrum* 42(5):345–346.

Díaz-Morales R (2015) Cross-device tracking: Matching devices and cookies. *Proc. IEEE Internat. Conf. on Data Mining Workshop* (IEEE, New York), 1699–1704.

Economides N, Lianos I (2021) Restrictions on privacy and exploitation in the digital economy: A market failure perspective. *J. Competition Law Econom.* 17(4):765–847.

Elliehausen GE (1998) The cost of bank regulation: A review of the evidence. United States Federal Reserve System, Staff Studies Series, Number 171, Washington, DC: Board of Governors of the Federal Reserve System.

Elliehausen GE, Kurtz RD (1988) Scale economies in compliance costs for federal consumer credit regulations. *J. Financial Service Res.* 1:147–159.

European Commission (2012) Impact Assessment accompanying the regulation of the European Parliament and of the Council on the Protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) and directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. SEC(2012) 72 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012SC0072&from=EN>.

European Commission (2021) Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM(2021) 206 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

European Data Protection Board (2019a) Guidelines 3/2018 on the territorial scope of the GDPR (Article 3), Version 2.1. [https://edpb.europa.eu/sites/default/files/files/file1/edpb\\_guidelines\\_3\\_2018\\_territorial\\_scope\\_after\\_public\\_consultation\\_en\\_1.pdf](https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_3_2018_territorial_scope_after_public_consultation_en_1.pdf).

European Data Protection Board (2019b) Opinion 5/2019 on the interplay between the eprivacy directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities. [https://edpb.europa.eu/sites/default/files/files/file1/201905\\_edpb\\_opinion\\_eprivacydir\\_gdpr\\_interplay\\_en\\_0.pdf](https://edpb.europa.eu/sites/default/files/files/file1/201905_edpb_opinion_eprivacydir_gdpr_interplay_en_0.pdf).

- European Data Protection Board (2021) Guidelines 07/2020 on the concepts of controller and processor in the GDPR, version 2.0. [https://edpb.europa.eu/system/files/2021-07/eppb\\_guidelines\\_202007\\_controllerprocessor\\_final\\_en.pdf](https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf).
- European Union (1995) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal, L*. 281:31–50.
- European Union (2002) Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications, lasted amended by Directive 2009/136/EC). *Official Journal, L* 201:37–47.
- European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/ec (General Data Protection Regulation). *Official Journal, L* 119:1–88.
- Farboodi M, Mihet R, Philippon T, Veldkamp L (2019) Big data and firm dynamics. *AEA Paper Proc.* 109:38–42.
- Frankenreiter J (2022) The missing “California effect” in data privacy law. *Yale J. Regulation*. Forthcoming.
- Gal MS, Aviv O (2020) The competitive effects of the GDPR. *J. Competitive Law Econom.* 16(3):349–391.
- Geradin D, Karanikoito T, Katsifis D (2021) GDPR myopia: How a well-intended regulation ended up favoring Google in ad tech. *Eur. Competition J.* 17(1):47–92.
- Gill P, Erramili V, Chaintreau A, Krishnamurthy B, Papagiannaki D (2013) Follow the money? Understanding economics of online aggregation and advertising. *Proc. ACM Internet Measurement Conf.* (ACM, New York), 141–148.
- Godinho de Matos M, Adjerid I (2021) Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Sci.*, ePub ahead of print September 8, <https://doi.org/10.1287/mnsc.2021.4054>.
- Goldberg S, Johnson G, Shriver S (2021) Regulating privacy online: An economic evaluation of the GDPR. Working paper, SSRN, submitted July 18, 2019; update March 29, 2021.
- Goldfarb A, Treffer D (2018) *Artificial Intelligence and International Trade* (University of Chicago Press, Chicago).
- Goldfarb A, Tucker CE (2011) Privacy regulation and online advertising. *Management Sci.* 57(1):57–71.
- Goldsmith J, Wu T (2006) *Who Controls the Internet? Illusions of a Borderless World* (Oxford University Press, Oxford, UK).
- Gopal RD, Hidaji H, Patterson RA, Rolland E, Zhdanov D (2018) How much to share with third parties? User privacy concerns and website dilemmas. *Management Inform. Systems Quart.* 42(1):143–164.
- Grover V, Cheon MJ, Teng JT (1996) The effect of service quality and partnership on the outsourcing of information systems functions. *J. Management Inform. Systems* 12(4):89–116.
- Hagiu A, Wright J (2020) Data-enabled learning, network effects and competitive advantage. Working paper. <https://app.scholarsite.io/julian-wright/articles/data-enabled-learning-network-effects-and-competitive-advantage-3>.
- Hu X, Sastry N (2019) Characterising third party cookie usage in the EU after GDPR. *Proc. 10th ACM Conf. on Web Sci* (ACM, New York), 137–141.
- Janssen R, Kesler R, Kummer M, Waldfogel J (2021) GDPR and the lost generation of innovative apps. NBER Economics of Digitization Workshop, March 19, 2021.
- Jia J, Jin GZ, Wagman L (2021) The short-run effects of the general data protection regulation on technology venture investment. *Marketing Sci.* 40(4):661–684.
- Jin GZ, Wagman L (2020) Big data at the crossroads of antitrust and consumer protection. *Inform. Econom. Policy* 54:100865.
- Johnson G, Shriver S, Goldberg S (2020) Privacy & market concentration: Intended & unintended consequences of the gdpr. Preprint, submitted November 15, 2019; last revised January 14, 2021, <https://dx.doi.org/10.2139/ssrn>.
- Jones CI, Tonetti C (2020) Nonrivalry and the economics of data. *Amer. Econom. Rev.* 110(9):2819–2858.
- Ke TT, Sudhir K (2020) Privacy rights and data security: GDPR and personal data driven markets. Preprint, submitted July 17, <https://dx.doi.org/10.2139/ssrn.3643979>.
- Lefrere V, Warberg L, Cheyre C, Marotta V, Acquisti A (2019) The impact of GDPR on Ad-supported content providers. Working paper, Workshop: Workshop on Information Systems and Economics WISE 2019, Munich, Germany.
- Lerner A, Kornfeld Simpson A, Kohno T, Roesner F (2016) Internet Jones and the raiders of the lost tracker: An archaeological study of web tracking from 1996 to 2016. Holz T, Savage S, eds. *Proc. 25th USENIX Security Sympos.* (USENIX Association), 997–1013.
- Libert T (2015) Exposing the hidden web: An analysis of third-party HTTP requests on one million websites. *Internat. J. Comm.* 9: 3544–3561.
- Libert T, Graves L, Nielsen RK (2018) Changes in third-party content on European news websites after GDPR. Technical report, Reuters Institute for the Study of Journalism, University of Oxford, Oxford, UK.
- Lim B, Wen HJ (2003) Web services: An analysis of the technology, its benefits, and implementation difficulties. *Inform. Systems Management* 20(2):49–57.
- Loh L, Venkatraman N (1992) Determinants of information technology outsourcing: A cross-sectional analysis. *J. Management Inform. Systems* 9(1):7–24.
- Rustad ML, Koenig TH (2019) Towards a global data privacy standard. *Florida Law Rev.* 71:365–453.
- Schelter S, Kunegis J (2018) On the ubiquity of web tracking: Insights from a billion-page web crawl. *J. Web Sci.* 4(4):53–66.
- Schmitt J, Miller KM, Skiera B (2020) The impact of privacy laws on online user behavior. Preprint, submitted March 16, 2021, <https://dx.doi.org/10.2139/ssrn.3774110>.
- Schwartz PM (2019) Global data privacy: The EU way. *New York University Law Rev.* 94:771–818.
- Schwartz PM, Peifer K-N (2017) Transatlantic data privacy law. *Georgetown Law J.* 106:115–179.
- Sharma P, Sun Y, Wagman L (2019) The differential effects of new privacy protections on publisher and advertiser profitability. Working paper, SSRN, submitted December 17, 2019; updated April 8, 2021.
- Solomos K, Ilija P, Ioannidis S, Kourtellis N (2019) Clash of the trackers: Measuring the evolution of the online tracking ecosystem. *Proc. Network Traffic Measurement Anal. Conf. (TMA'20), Berlin, Germany* (IFIP).
- Sørensen J, Kosta S (2019) Before and after GDPR: The changes in third party presence at public and private European websites. *Proc. World Wide Web Conf* (ACM, New York), 1590–1600.
- Tobin O (2019) GDPR: 3 areas of ambiguity. *Privacy Data Protection* 20(2):15–16.
- Urban T, Tatang D, Degeling M, Holz T, Pohlmann N (2020) Measuring the impact of the GDPR on data sharing in ad networks. *Proc. 15th ACM ASIA Conf. Comput. and Comm. Security* (ACM, New York), 222–235.

- Vogel D (1995) *Trading Up: Consumer and Environmental Regulation in a Global Economy* (Harvard University Press, Cambridge, MA).
- Watjatrakul B (2005) Determinants of is sourcing decisions: A comparative study of transaction cost theory versus the resource-based view. *J. Strategic Inform. Systems* 14(4): 389–415.
- WhotracksMe (2018) GDPR: What happened? Technical report.
- Willcocks L, Oshri I, Hindle J (2010) *To Bundle or Not to Bundle? Effective Decision-Making for Business and Its Services* (LSE/Accenture Whitepaper).
- Yoo Y, Boland RJ Jr, Lyytinen K, Majchrzak A (2012) Organizing for innovation in the digitized world. *Organ. Sci.* 23(5):1398–1408.
- Zhuo R, Huffaker B, Claffy KC, Greenstein S (2021) The impact of the general data protection regulation on internet interconnection. *Telecomm. Policy* 45(2):102083.