SIIA POLICY

Adopting a Risk-Based Approach to the Regulation of Data Practices

This content reflects the view of the Software & Information Industry Association (SIIA), and may not reflect the individual views of each SIIA member company.

What do efforts to combat human trafficking have in common with schemes intended to defraud elderly Americans and those with mental health conditions? The short answer is they rely on the ability to use data compiled and analyzed by commercial publishers. Yet in discussions taking place today in Washington, DC and in state capitols across the country, these entities are lumped together with a host of others—many essential to advancing critical societal needs, and others that undermine individual rights and social cohesion—under the pejorative term “data broker.”

That term is politically clever, but practically useless. What these policymakers are doing in reality is regulating not entities, but an activity: commercial publishing. The purpose of this white paper is to provide nuance to the discussion around that activity. We believe regulation of commercial publishing must be layered and tailored based on the risk created by the kind of information being used and the nature of the harm that could flow from that use. Armed with this more nuanced view, we then hope to illustrate the diversity of legal, practical and ethical considerations necessary for fair and productive regulation of the industry as a whole.

The Benefits of Data and the Challenges with One-Size-Fits-All Restrictions

"Data broker" is a misnomer. Many so-called data brokers are actually commercial publishers. They routinely help both businesses and governments deter, prevent, and track down fraudulent criminal activity. Publishers disseminate information ranging from business-to-business news, to curricula about the Civil War, to securities pricing, to databases of case law and other public records, to scientific, technical and medical articles. These publications are used for purposes ranging from academic research to corporate due diligence. Taken as a whole, these products and services can enable commerce, prevent crime, and provide the building blocks of ideas – the backbone of functioning markets and a functioning democracy.

While laudable efforts are underway to limit the use and publication of personal data in ways that cause harm to individuals, the positive, societally beneficial practices that rely on data are treated identically in a growing discourse committed to stopping "data brokers." For all the pithiness of this phrase, these legislators misunderstand that they are actually regulating an activity—commercial publishing—that is both valuable to society and constitutionally protected. Indeed, the breadth of the definitions that undergird virtually all efforts to regulate "data brokers" could cover everything from a database of news articles to a phone book, and we submit that it is fundamentally unhelpful in making policy choices.[1]

Recognizing that legislatures are regulating publishing will lead to more nuanced approaches that both preserve free speech rights and protect societally valued activity. The one-size-fits-all approach that has dominated policy discourse for years is misguided and, if it continues unabated, will lead to federal and state law that will directly harm consumers, undermine confidence in the U.S. economy, weaken critical civil rights protections, and restrict government functions in essential ways that we take for granted.

1 Efforts to regulate the collection, sale, and use of data on U.S. individuals almost uniformly focus on the entities involved in these activities and do not distinguish among the types of information involved and the ways in which the information is used. See, e.g., Health and Location Data Protection Act ("The term 'data broker' means a person that collects, buys, licenses, or infers data about individuals and then sells, licenses, or trades that data."); California SB 362 ("DELETE" Act) ("'Data broker' means a business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship.").

Consider, for example, the numerous activities that the information industry enables:

- **Law enforcement.** Federal and local law enforcement agencies rely on "data brokers" to locate suspects, victims, and witnesses to crimes. Even agencies such as the Federal Bureau of Investigation rely on products provided by "data brokers" because they enable "FBI investigative personnel to perform searches from computer workstations and eliminates the need to perform more time-consuming manual searches of federal, state, and local records systems, libraries, and other information sources."[2] These private-sector tools can also organize and analyze publicly-available data, infer patterns, and efficiently sort relevant information from irrelevant information. The government routinely uses these tools in investigations of large-scale, complex, and sometimes international criminal schemes that use shell companies to avoid detection. They have been used to investigate, for example, human trafficking and fentanyl distribution networks.

- **Combating money laundering, corruption, and terrorism.** Financial institutions and other businesses rely on both non-public and publicly available data sources to help them meet "know your customer," anti-money laundering, anti-terrorism, and anti-human trafficking obligations, and to comply with other financial laws, regulations, and industry practices. For example, banks have specific obligations when opening accounts for Politically Exposed Persons ("PEP") who are close relatives of senior government officials. PEP lists are compiled from publicly available media combined with non-public data. These services enable the implementation of best practices in line with international objectives for corporate governance and efforts to combat bribery and corruption around the world.

- **Child support enforcement.** State and local agencies use data sets provided by publishers to locate individuals who are delinquent in paying their child support obligations. The Association for Children for Enforcement of Support reports that public record information provided through commercial vendors helped locate over 75 percent of the "deadbeat parents" they sought.[3]

2 Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations for Fiscal Year 2000: Hearings on H.R. 2670/S.1217 Before Subcomm. for the Depts of Commerce, Justice, and State, the Judiciary, and Related Agencies of the S. Comm. on Appropriations, 106th Cong. 280 (1999).
3 Comments of Gail H. Littlejohn, Vice President, Gov't Affairs, & Steven M. Emmert, Dir., Gov't Affairs, Reed Elsevier Inc., LEXIS-NEXIS Group (Mar. 31, 2000), available at http:// www.sec.gov/rules/proposed/s70600/littlej1.htm; see also Financial Information Privacy Act: Hearings on H.R. 4321 Before the H. Comm. on Banking and Financial Services, 105th Cong. 100 (1998) (statement of Robert Glass).

- **Insurance.** Insurers of all shapes and sizes access such data each day to underwrite policies and pay claims.

- **Product safety.** Companies use this information to provide consumers and auto dealers with a vehicle's accident history, alerting consumers to whether they are potentially buying a "lemon," and to put both dealers and consumers on notice that the vehicle is subject to a safety recall.

- **Tax Compliance and prevention of waste and abuse.** Governments use databases of real estate records to detect tax avoidance. Moreover, they use our members' tools to investigate potential abuse of public benefits programs, such as pandemic payment fraud, by using inferences from a combination of public real estate records, social media posts, and other public sources.

- **Fraud prevention.** Identity theft and other forms of fraud are a constant threat to consumers and businesses alike. Companies often leverage public record data to authenticate consumers in order to prevent identity theft and fraud. This can include an insurance company obtaining data from the DMV—or a vendor reselling this data—in order to authenticate a consumer's true identity and risks. It can also include asking "out of wallet" questions, which are those a fraudster would be unlikely to know, such as "Which of the following five addresses is a past home address of yours?" or "Which of the following cars did you once own?" The answers to these questions can be found in state real property records or publicly-available Uniform Commercial Code filings.

- **News-gathering and publishing.** Newspaper companies regularly obtain the information used in brokered data products to report on crimes, detect possible corruption or conflicts of interest, and publish stories involving the operation or safety of motor vehicles.

To be sure, improper data practices exist, and certain practices flagged by critics may be worthy of opprobrium. These practices, however, have nothing whatsoever to do with the status of the publisher as a "data broker," but rather with the nature of the informational injury suffered by the consumer. And these injuries can be quite different: the harm suffered by a consumer from an incorrect credit report (defamation) differs entirely from that caused by the unauthorized disclosure of a cancer diagnosis (public disclosure of private fact). Use of the term "data broker" has been effective as a pejorative, but has resulted in overbroad and ill-advised policy proposals.

A New Approach: Towards a Risk-Based Framework

The preservation and advancement of socially beneficial uses of data requires a regulatory approach that addresses informational injury based on the risk of harm stemming from its use. Under a risk-based approach, regulation can then be applied specifically to those uses that present the highest risks to consumers, security or even the national interest, and avoid wrapping in the productive aggregation and publishing of societally or commercially useful information.

Much of the legislation in Europe and even in the United States rests on the premise that personal data should be regulated as property. Europe's GDPR, for example, as well as many state privacy laws in the U.S., invokes this framework, which grants "owners" of this data control over its use. This is evident, for example, in the assumptively non-permissive "opt-in" approach taken under the GDPR and for "sensitive data" in the United States – but also even the "opt-out" approach taken for mere personal data under every U.S. state privacy law. The mere inclusion of personal information without consent constitutes a "trespass" on the data subject's rights. Furthermore, at the time of writing, Congress is considering some bills applying to Section 702 of FISA that would apply a rigid rule against data broker purchases. This is not programmatic legislation, so Congress cannot simply revert to prior language if this property-based framework creates unintended consequences in a national security context. Instead, getting something like this wrong can have significant and irreversible effects.

We believe this quasi-property approach is misguided and will not work for at least three reasons. First, property rights in personal data do not exist. Property rights rest "on an assumption that the rights-holder has superior knowledge about the best uses of the property, [and] would know when to exclude, when to share, and when to sell the property, and would do so without causing significant problems for others."[4] The First Amendment views the regulation of information and publishing as a last resort, not a first one, and it recognizes that the dissemination of certain types of facts about individuals is not an injury. It does not require consent. For example, publicly-available information is protected by constitutional design, and does not implicate privacy concerns once released and widely available.

4 Jane R. Bambauer, How to Get the Property out of Privacy Law (2023).
5 See Sorrell v. IMS Health, Inc., 564 U.S. 552 (2011) (holding that restrictions on commercial speech, particularly content and speaker-based restrictions, are subject to heightened intermediate scrutiny).

Although the jurisprudence on commercially-available information is less clear, it also retains significant protections, especially, as is often the case, when specific speakers or content is restricted.[5] There will be in many cases no injury flowing from the dissemination of accurate public domain data.

Second, property rights are designed to prevent a tragedy of the commons: when the property owner is forced to internalize the cost of use, they tend to make better decisions. Information, however, is a public good. The consumer is not in a position to balance the contravening but valid interests of society (anti-fraud technologies and the prevention of money laundering), data controllers and processors (personalized speech), and consumers themselves (location matching). A property-based approach then risks eliminating pools of societally beneficial data with little or no corresponding benefit to consumers – or even harms unforeseen at the time of an opt-out or deletion request.

Third, the grant of an exclusive right in personal data ignores the fact that privacy rights change over time and in response to technology. What a reasonable person might find to be a "trespass" in 1987 would look very different today.

These three factors suggest that the regulation of commercial publication would benefit from a risk-based approach built around its uses – not the act of dissemination or even the underlying data itself. This aligns more with an understanding of privacy in the context of civil torts and the common law's historical development of different classes of informational injury. Such an approach permits greater flexibility for policymakers to balance competing interests around both data uses and data privacy, without the overbreadth that a property approach guarantees.[6] At a high level, the factors that define the benefits and risks surrounding published data are: (a) the type of data collected and methods of collection, and (b) the data's end uses.

Commercial Publishing Presumptively Receives Constitutional Protection

The first question that a regulator ought to consider is whether the data it seeks to regulate is in the constitutionally protected public domain. The Supreme Court has made clear that "the creation and dissemination of

5 See Sorrell v. IMS Health, Inc., 564 U.S. 552 (2011) (holding that restrictions on commercial speech, particularly content and speaker-based restrictions, are subject to heightened intermediate scrutiny).
6 See Bambauer (2023).
7 Sorrell v. IMS Health Inc., 564 U.S. 552, 570 (2011).

information is speech for First Amendment purposes."[7] The State may not infringe these rights to protect a generalized interest in consumer privacy, as such restrictions burden both the businesses whose speech they restrict and the users of the information who are entitled to receive it.[8] The constitutionally protected public domain also consists not only of information released by the government, but that which is widely available in private hands – as well as opinions or inferences drawn from that information. While legislators can prohibit the use of public domain information (for example, by adding elements to a stalking offense), regulating the publication of public domain information is almost certainly unconstitutional, as such a statute will likely have problems with vagueness, overbreadth, discriminating among speakers, or content – all of which are likely facial violations of the First Amendment.[9]

The second category of information we will call "commercially available information." This category of information has two key characteristics: (1) it is not available to the public but made available by publishers for commercial use; and (2) it may be acquired in circumstances in which the consumer has an expectation of privacy. Misuse of this kind of information could cause interference with historically recognized privacy interests by revealing a private fact, the disclosure of which would be offensive to a reasonable person. Examples of such information include video sales rental data, geolocation data, or a consumer's purchase history.[10] Not all disclosures of such information are offensive, and a productive policy debate can be had about the circumstances under which such information should be used. In the same vein, inferences made from a mixture of public information and this second category of commercially-available information would be subject to a lower level of scrutiny.

Finally, there is a category of highly-sensitive information that experience has taught creates an immediate high risk of an informational injury. In the common law, certain kinds of statements remain defamation per se, but injury in this category is not limited to the classic common-law privacy torts. In the more modern era, higher risk disclosures include unauthorized disclosures of social security or drivers' license numbers, financial account information, and similar kinds of data that, when unlawfully disclosed, readily leads to identity theft.[11]

---

7 Sorrell v. IMS Health Inc., 564 U.S. 552, 570 (2011).
8 See generally E. Volokh, Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking About You, 52 Stan. L. Rev. 1049, 1081 (2000); Stanley v. Georgia, 394 U.S. 557, 564 (1969).
9 See Rosenberger v. Rector & Visitors of Univ. of Va., 515 U.S. 819, 828 (1995) ("In the realm of private speech or expression, government regulation may not favor one speaker over another.").
10 See Video Privacy Protection Act, 18 U.S.C. § 2710 (1988).
11 See Driver's Privacy Protection Act, 18 U.S.C. § 2721 (1997).

Some Information Functions Present Little or No Consumer Privacy Risk

The intended and actual end use of published information directly implicates the risk involved in this business model. Commercial publishing advances many positive purposes largely taken for granted by both consumers and businesses. From the perspective of evaluating and mitigating risk, these functions can be divided into (a) "white hat" activities widely viewed as beneficial to society; (b) commercial purposes that enhance efficiencies and help businesses reach consumers, but may implicate privacy concerns; and (c) "black hat" uses that enable fraud by private entities or even adversarial foreign governments.

- **White hat information functions.** "White hat" functions are those that gather and make productive uses of data for societally-beneficial purposes. Gathering and assessing data to help prevent fraud or assist law enforcement investigations fall into this category. While policymakers may often disagree on politically sensitive topics—for example, law enforcement priorities or civil liberties protections for citizens—publishers that merely assist police with investigations are well outside the appropriate focus for these debates.

- **Commercial information functions.** This type of targeted information sharing typically involves providing commercial data sets to third parties that may use information for a myriad of legitimate purposes. These purposes may include personalizing user interfaces and experiences, targeted advertising, or crafting marketing messages. Although advertising and marketing has become the focus of several state privacy laws in recent years, it is far from a social ill. Appropriate regulatory responses may focus on preventing abuses as well as incentivizing the use of privacy enhancing technologies to mitigate the risk of misuse of personal information.

- **Black hat information functions.** Black hat functions are those that include mining personal data and selling it to fraudsters and other malicious actors, and may be targeted toward such buyers. This data may be CAI, commercial speech, or even aggregating public data for nefarious purposes. Of course, the use-based risk is likely worse if this data is nonpublic or sensitive. These entities are the most oft-cited example of brokers in headlines and discussions around the need for regulation – however, they represent only a sliver of the industry as a whole.

A Risk-Based Framework is Appropriate for Commercial Publishing

A risk-based analysis requires an assessment of the type of information involved, as well as the use of that information. Following this approach, there should be a presumption that publishing of commercially available information and public information for a white hat function is permissible. Among others, these uses could include anti-fraud efforts, technologies designed to protect the health or safety of a consumer, or data uses mandated by existing law such as compliance with judicial orders or regulatory investigations. This aligns with the conception of privacy as a tort. Because these data uses do not cause substantial injury–and in fact benefit consumers and society at large–they should not be restricted by law or a rigid framework of consumers possessing exclusive domain over information whose dissemination benefits large networks of stakeholders. White hat uses of highly-sensitive data will require additional scrutiny.

Despite controversies around certain advertising practices, commercial purposes are also often beneficial. Consumer privacy interests may benefit from regulation in cases of particularly sensitive data. Yet this often has less to do with the data itself than its intended purpose: ostensibly "sensitive" data may be leveraged for routine or inoffensive commercial purposes, whereas even public domain information may be used to predict or maliciously target consumers. For example, many anti-fraud technologies screen the same data that would otherwise be co-opted for malicious purposes to protect vulnerable consumers.

Lastly, black hat uses that promote fraud or more easily disseminate private data to malicious state actors should be wholly prohibited. Furthermore, even if a publisher intends data provided to third parties to be used for societal or positive commercial purposes, this is not guaranteed once it is transferred. Therefore, contractual requirements placed upon third parties that flow down with this data, as well as routine audits confirming how the transferred information is being used–as is current practice within many publishers–can further mitigate use case risk and enhance consumer privacy interests.

Information Stewardship May Further Reduce Dissemination Risks

Lastly, a risk-based approach should account for an entity's stewardship of information. Regulators should explore ways to encourage if not require responsible data practices by all entities, tailored to the type of data involved and the prospective uses of that data. Data breaches are already regulated under existing law. Yet due to consumer privacy concerns distinct from outright breaches, many states have proposed to provide a safe harbor for companies whose cybersecurity programs comply with the NIST Framework.[12] These laws both incentivize and reward those companies that invest in agreed-upon cybersecurity protocols, and provide a roadmap for regulatory treatment of publishers.

Conclusion

As a rule, societally-productive uses of data as well as commercial and advertising uses would benefit from regulatory clarity. That clarity, however, requires a nuanced approach to regulating commercial publishing that distinguishes among the types and uses of data within the information ecosystem. Publicly-available data, for example, cannot and should not be restricted or risk running afoul of the First Amendment.

Conversely, the end uses of the data also provide a compelling argument for differentiating among commercial publishers. There will be no free lunch: for example, prohibiting the use of commercially available information by law enforcement will increase consumer privacy writ large, but will also carry with it the cost of criminals escaping. Furthermore, from a risk-based perspective, if data is misused in a privacy-invasive manner, we suggest regulating this intrusive use itself rather than attempting to solve the problem by expanding existing restrictions into new categories of data. The risk stems from the use of this data, not its availability or dissemination writ large. In fact, the use of this same data elsewhere may even reduce privacy risk.

12 Sorrell v. IMS Health Inc., 564 U.S. 552, 570 (2011). 8 See generally E. Volokh, Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking About You, 52 Stan. L. Rev. 1049, 1081 (2000); Stanley v. Georgia, 394 U.S. 557, 564 (1969). 9 See Rosenberger v. Rector & Visitors of Univ. of Va., 515 U.S. 819, 828 (1995) ("In the realm of private speech or expression, government regulation may not favor one speaker over another."). 10 See Video Privacy Protection Act, 18 U.S.C. § 2710 (1988). 11 See Driver's Privacy Protection Act, 18 U.S.C. § 2721 (1997).

At a high level, prohibiting commercial publishing or the collection of certain categories of data based on an undifferentiated fear of commercial publication, without the context of its use, would be a grave mistake. It would merely promote needless balkanization of useful information and likely harm, not enhance, consumer privacy or security.

ABOUT SIIA

SIIA is the principal trade association for the software and digital information industries. Our members include over 450 companies reflecting the broad and diverse landscape of digital content providers and users in academic publishing, education technology, and financial information, along with creators of software and platforms used by millions worldwide, and companies specializing in data analytics and information services. SIIA is the only association representing both those who develop and deploy AI engines and those who create the information that feeds environments.

To learn more visit www.siia.net/policy/

SIIA
POLICY