

Personal Data as an Environmental Hazard

Author : Ronen Avraham

Date : November 14, 2018

Omri Ben-Shahar, "Data Pollution," University of Chicago Public Law & Legal Theory Paper Series, No. 679 (forthcoming 2018), available at [SSRN](#).

What was the nature of the harm when data on 143 million Equifax consumers was stolen? More generally, what is the problem with personal data use and misuse by commercial players? The most immediate answer: privacy, individuals' privacy interests are infringed. But then the question becomes what is the problem with infringing one's privacy? Here, the answer usually is that infringing one's privacy infringes upon her autonomy, dignity, emotional wellbeing, and such. To these non-monetary harms, one can add various monetary harms such as monetary losses associated with identity theft and other economic losses. These personal harms have led privacy scholars to focus on the private and personal aspects of data breaches. This, in turn, has naturally also led them to focus on private law solutions, such as tort and contract law-type protections for individual's privacy interests.

However, despite years of attempting to combat irresponsible data sharing and handling, the problem persists. People treat their private personal information as if they do not much care about it. They may trade it quid pro quo for access to various services from navigation and communication services on their cell phones to participation in social networks, or even just to play Fortnite. And yet, when asked about how important privacy is to them, people overwhelmingly claim it matters a lot. Similarly, when people sue for damages for data breaches, they claim they suffered significant losses. The gap between what people claim (privacy matters) and what they do (sell it cheaply) is the so called "privacy paradox." How can this paradox be resolved?

Some scholars have suggested that the gap stems from the fact that people are misinformed, irrational, or rationally misinformed or even rationally irrational. In a brand-new paper titled "[Data Pollution](#)," Omri Ben Shahar offers a different solution.

In "Data Pollution," Ben Shahar shifts the focus of data breaches from the private harm suffered by individuals whose information was used to the societal harm writ large. And the harm to society Ben Shahar has in mind is not at all the aggregate harm of the class of individuals harmed, nor is it an abstract, derivative harm stemming from the emerging distrust of individuals in private or public institutions. Rather, it is a direct and concrete harm to the public ecosystem. Indeed, the social harm might occur even when individuals suffer no private harm, or even benefit from willingly sharing their private data.

How can it be? Think about the potential harm to the integrity of the American election system as a result of Facebook data-share with Cambridge Analytica. The individuals whose data was shared might be personally happy about the data share, yet the truly troubling problem was to the American democratic ecosystem. Or think about users of the Strava fitness app who share their running trails with the world, not realizing they were revealing locations of secret military bases, and therefore harming national security interests. This type of harm, Ben Shahar argues, stems from the fact that data is hazardous and if not handled well, might create dangerous "data pollution."

When data misuse is viewed less as a personal harm to the privacy interest of individuals and more as a

social harm to the public sphere, the paradigm shifts. Thus, according to Ben-Shahar, policy makers' focus should not only be on the best ways to protect individuals from commercial players in the public sphere, but also on the best ways to protect the public sphere from individuals sharing their personal data with commercial players. The problem, in other words, is not just that commercial players trade individuals' private data without adequately compensating them in an ex-ante user agreement, a fact that can be explained by individuals being misinformed or irrational. Nor is the problem just that commercial players mishandle individuals' private data without adequately compensating them in an ex-post tort suit, a fact that can be explained by the difficulty in proving causation, estimating the harm, etc. Rather, according to Ben Shahar, the problem also is that private individuals trade their own private data at a price which does not reflect the negative externalities they create. It is this last feature – the negative externality – which is Ben Shahar's main contribution to the literature and which will be my focus here.

“Data are to this century what oil was to the last one.” This quote opens Omri Ben-Shahar's new thought-provoking paper. Data is the fuel of the information economy and like fuel in the oil economy, data pollutes, and it pollutes the digital ecosystem in ways which directly disrupt the public interest. For years, privacy advocates have spent much energy unsuccessfully trying to raise people's awareness to the privacy interest in their own data, while not spending any energy raising people's awareness to the social problem of data pollution. And yet, the external social costs associated with shared data might be, sometimes at least, much larger than the private costs.

It is as if policy makers had warned people about the fire hazards from using household kerosene lamps and had not warned them about the contribution to global heat from black carbon emissions. Only that in the data pollution case, unlike in the kerosene case, the social harm from the emission might be much more significant and more dangerous than the private harm to any individual.

The external costs of data emission are neglected any time individuals agree to various user agreements; thus, causing the sale of the private data to be at a price which is lower than the socially-optimal price. The external costs of data emission are also neglected when individuals demand compensation after a privacy breach, thus causing the wrongdoer to pay compensation which is lower than the socially-optimal level of compensation. As a result of these two problems, the level of information which is shared by individuals is excessive, just like pollution. The analogy to pollution and externality is what helps us better see the problem as it really is.

The analogy to pollution is of course not one-to-one; there are several differences between data pollution and industrial pollution. For example, there are two major types of externalities in data pollution, only one of which also clearly exists in industrial pollution. The first type is similar to the harm in industrial pollution: shared data can be aggregated and used or misused in ways which affect the public interest at large. The second type of externality is on other users. This type is unique to data pollution—individuals often share information not just about themselves, but also about others who do not want their information to be shared. An example would be when individuals agree to share their contacts. Perhaps the closest analogy from “real” pollution here is second-hand smoking, where people ignore the costs they incur to friends and family members near them.

Another difference is that data pollution does not only create negative externalities, it also creates positive externalities (think about predicting flu epidemics by looking at google pharmacy searches). Indeed, data collection is among the most productive activities of the 21st century.

Still, the analogy is helpful, and not just in providing a fresh and helpful re-conceptualization of data misuse problems. The analogy assists in analysing potential policy solutions. Data misuse can be regulated by policies long used to control industrial pollution, such as emission quotas, Pigouvian taxes,

and legal liability. Indeed, one should not be surprised that private law tools, such as contract and torts, were unable to control the social problem of data pollution. After all, they failed to control industrial pollution as well. And they failed, Ben Shahaar argues, primarily because they cannot handle *externalities* well.

The analogy to industrial hazards pertains not just at an abstract level, but also in specific details. What is the equivalent of quotas imposed on polluters? Restrict what data can be collected, from whom, by whom, for what purposes and for how long. Sounds crazy, right? But this is exactly what European regulators do. What about a Pigovian tax? Ben Shahaar proposes, contra to current practice and scholarship, to *tax individuals* who share personal information quid pro quo for various services. No more free access to cable TV in return for allowing cable companies to collect data on subscribers. A tax would make subscribers (and cable companies) internalize the social external cost associated with the emission to the digital sphere of subscribers' personal data. And, what about liability for data breaches? Because, unlike physical spills, data spills cannot be cleaned up ex-post. Ben-Shahaar proposes imposing liability which equals the *expected* social costs of the spill, hopefully generating enough deterrence to prevent these spills from occurring in the first place.

Occasionally, the analogy can only be pushed so far. For example, there is no equivalent, yet, in data pollution to cap-and-trade in industrial pollution. Since the objective is to prevent massive data from being accumulated in the hands of a few players, policy makers should forbid players from trading data with each other. Rather than cap-and-trade, policy makers should enforce a strict cap-and-don't-trade policy. Or, alternatively, since accumulation of data in the hands of the few is analogous to the problem of "hot-spots" in industrial pollution, there is even more that can be borrowed from the regulation of industrial pollution to the regulation of data pollution.

What is Ben Shahaar's preferred tool for combating data pollution? Readers who want to know that are left to find the answer themselves. Instead, in the space left for me here, I will make two comments.

The first comment has to do with the insight that private individuals are themselves not just victims, but also wrongdoers—they are polluters. True, the magnitude of the harm they can create is much smaller than the potential harm of data aggregators, but still... they are polluters. Recognizing that, recall that Ben Shahaar proposes a tax on individuals who data-share. But taking the idea of end-user pollution seriously brings to mind another solution: end-user liability.

Consider a case of an individual who clicks on a malicious link, which enables hackers to install ransomware on his contacts' computers. As a result, some of his contacts need to pay thousands of dollars to free their computers. The law usually will not find the individual liable. But why not? One potential answer is that end-users lack the expertise to protect their computers effectively as well as lack the resources to pay damages. What is missing from this answer is the availability of insurance, perhaps even mandatory insurance, for all online users. Insurance might help individuals protect themselves and others not just by providing coverage, but also by offering technical means to prevent the losses before they occur as well as the technical help to mitigate them after they have occurred. With insurance, it is no longer clear why end-user liability is not currently an option.¹

My second comment tries to push the pollution metaphor even further from the context of data into other contexts. In a recent working paper, my co-authors and I used the pollution metaphor in the context of civil procedure.² Yet here, I would like to use it in the context of discrimination.³

Consider a minority religious student group, which organizes an event on university grounds. The group announces that seating will be separated by gender. Should the university approve the event under these terms? This is not a hypothetical example, but it is based on real events that happened recently in

the U.K. On the one hand, gender-separated events infringe upon notions of equality as by now we are accustomed to thinking that separate but equal is usually not equal.⁴ On the other hand, what if most members of the student group (both men and women) prefer gender-separated seating? Should we not respect their preferences?

Policy makers who want to forbid gender-separated seating can do it in one of two ways. First, a paternalist approach, where they basically ignore the preferences of the group members under the assumption that their consent to separated seating is not free, well informed or rational. Second, policy makers may choose to protect a minority-within-the-minority, which prefers gender-mix seating but is subject to social pressures to comply with the more extremist gender-separated agenda. In the U.K., the Equality and Human Rights Commission has forbidden such events (with some exemptions for religious prayers) exactly on these grounds.⁵

Both ways are problematic. Being paternalistic towards group members by arguing they are misinformed, irrational or not free is problematic exactly because it is...paternalistic. And, defending the minority-within-the-minority is problematic because it is not always clear that there is such a group at all. And yet, our intuition many times is to forbid such events.

Adopting the pollution approach to discrimination exposes the problem with having segregated events on university grounds even when the group members want it. It reveals that "the toxicity from discriminatory treatment degrades the environment also for those not discriminated." (P. 8.) Segregation (on racial, gender or other forbidden grounds) pollutes the university grounds for the rest of the community. It is akin to a public "moral nuisance."

Ben Shahaar's "Data Pollution" is an insightful paper. Every time I thought Ben Shahaar could not possibly push the analogy to environmental hazards further, I discovered I was wrong. It remains to be seen if others will find it a useful conceptualization of the problems associated with the use and misuse of personal data by commercial players as much as I did.

1. Ronen Avraham & Joachim Mayer, *End User Liability* (On file with author).
2. Ronen Avraham, William H.J. Hubbard, & Itay E. Lipschitz, [Procedural Flexibility in Three Dimensions](#) (Apr. 27, 2018).
3. Ben Shahaar shortly discusses discrimination in his paper, but it is in a different context than the one I will be using here.
4. Since we are not talking about a religious practice such as a prayer there are no issues of freedom of religion involved.
5. See UK Equality and Human Rights Commission, [Gender Segregation at Events and Meetings](#).

Cite as: Ronen Avraham, *Personal Data as an Environmental Hazard*, JOTWELL (November 14, 2018) (reviewing Omri Ben-Shahaar, "Data Pollution," University of Chicago Public Law & Legal Theory Paper Series, No. 679 (forthcoming 2018), available at SSRN), <https://torts.jotwell.com/personal-data-as-an-environmental-hazard/>.