

**REMARKS BY SALOMÉ VILJOEN**

<https://doi.org/10.1017/amp.2023.63>

We—people who socialize, shop, and work online—are generating a lot of data. Since the advent of computers to 2017, the sum of all the data produced was about ninety-four zettabytes (ninety-four billion gigabytes). We are expected to produce more than that in just this year alone. This data creates a lot of value, but its collection has also caused growing concerns about how it is being used to surveil people, manipulate their behavior, and expose them to harm.

There is general consensus among scholars of law and technology that existing data privacy laws are inadequate to meet this moment, and that something must be done. But what exactly, is an ongoing and lively debate. I argue that this debate proceeds from an incorrect descriptive account of the data economy. Debates over how to reform or regulate privacy typically focus on the “vertical relation” between a user, like myself, and a data platform, like Google, and how the law governs our exchange: data from me for services from them. Focus on this vertical relation generates a related set of questions: what kind of data Google should be able to collect about me; how much value each party gets from the exchange; whether I truly consented to data collection; and the scope of what that consent covers (e.g., whether it is exhausted at the point of exchange, or whether I retain residual consent rights over downstream uses of that data that may affect me).

However, focusing on this vertical relation alone is an incomplete account of how the digital economy actually works to produce value (and harm) from social data production. When data is collected from a user, they are also placed into “horizontal relation” with other people. This horizontal relation, or data relation, reflects the notion that data collected from one person can almost never be said to be truly just about that person. Instead, data is almost always, in at least some fashion, also about others. And this feature—our data relations—is a lot of what makes data collection worthwhile for companies and other actors in the digital economy.

This observation is simple and straightforward (and I am far from the first person to make it), but its significance for data governance law is overlooked. Our data relations speak to, express, or capture the basic fact of our sociality. We are like one another, and therefore data about people like me—who share certain characteristics with me—can reveal meaningful things about me, too. These revelations are economically and socially useful, and in some sense, they have always been economically and socially useful. But it is only due to recent improvements in our technological capacity—vastly improved processing power, new data science and machine learning methods—that their widespread exploitation (in the descriptive, not pejorative sense) is economically and socially *feasible*. It is only relatively recently that the ubiquity and scale of digitally mediated human behavior allows companies and others to apprehend, make sense of, and act on the myriad ways we relate to (and can be related to) one another.

\* \* \* \*

Let us consider an example. Anna is four weeks pregnant. Ever since she began trying to become pregnant two years ago, Anna uses Apricot, a popular fertility tracking app and social media site. Anna uploads information on her ovulation cycle, eating habits, exercise patterns, sleep, mood changes, and a range of other physical symptoms. Apricot aggregates and synthesizes all this data that users share—not to sell people’s data directly, which they reassure users they would never do, but to sell insights about their user base to clients such as advertisers, employment consultancies, consumer credit agencies, and others.

All of this makes Apricot a lot of money! After all, early pregnancy data is very valuable. Having a child is a big event in a consumer’s life; this is when Anna is likely to change brand loyalties and purchasing behaviors, so advertisers are keen to be the first to get to her in this state of flux. But this

lucrative time is also a risky one. The legal status of Anna's first trimester pregnancy and her decisions about whether to stay pregnant are personally and legally fraught.

First trimester pregnancy is a vulnerable time in other ways too. Potential or current employers, creditors, or insurers might also be interested in knowing who is about to experience an expensive and disruptive life event (and who on the other hand is likely to be a reliable, healthy, productive worker). This is not to say that any of these entities want to specifically target or identify pregnant people. However, the autonomous systems increasingly used in these sectors to sift and identify people are designed to optimize for low-cost, low-risk people and are thus primed to identify and exclude the pregnant, even if wholly unintentionally.

There are two general types of response regarding how this data flow from Anna ought to be governed. The first cluster of approaches I refer to as the "propertarian approach" to data governance. The basic intuition here is that Apricot is getting rich from this valuable resource that Anna is providing for free, and that this arrangement is one of unjust enrichment. This is Anna's data, and she deserves to be paid for it! In response, propertarians tend to propose a labor or property right for Anna in her data, so that she can command a fair price in the market for her data.

The second cluster of approaches I refer to as the "dignitarian approach" to data governance. Dignitarians would argue that payment ignores the original sin of data extraction: the unjust violation of this inner sphere of Anna's life. Pregnancy and fertility status is very personal and intimate, and to gain access to knowledge about it, and render this knowledge about Anna legible to any number of systems, violates Anna's dignity, her ability to enjoy a privileged relationship to that information, and her right to determine who knows it on what terms. Moreover, on this view, knowledge of Anna's fertility status puts her at risk of being manipulated, and can thus be used in ways that undermine and thwart her ability to act autonomously. In response, dignitarians may advocate for granting Anna inalienable, or fundamental rights to her data that do not extinguish at the point of collection. Dignitarians also advocate for much higher thresholds for what counts as consent from Anna.

Let us imagine we adopt one of these proposals. In the propertarian instance, we grant Anna a wage right in her data, and she secures payment from Apricot for her data. In the dignitarian instance, we impose robust consent requirements on the terms of collection, and impose a requirement on Apricot that they not use Anna's data in any way that may adversely affect her. Either would seem to secure a real improvement along the vertical relation between Anna and Apricot.

However, both proposals leave unaddressed the horizontal relation between Anna and others like her—let us call one of them Becca. To understand this data relation between Anna and Becca, we need to understand a bit more about how Apricot uses data collection in its business. Anna uploads information regarding her fertility and pregnancy to Apricot, who sells insights about their user base (their behavioral patterns and trends) to clients of theirs. These entities, in turn, combine these insights with other data on users' television viewing and online purchasing patterns, to try to pick up on early indicators, the behavioral signs and patterns, of pregnancy (that valuable time in a consumer's life).

Becca, who does not use Apricot, fits the behavioral patterns that emerge from the analysis. Like Anna and others, she exhibits very similar patterns of changed television viewing and online purchasing behavior, indicating that she too is in the first trimester of a pregnancy. And indeed, a prospective employer, relying on the services of an autonomous candidate profile sorter, flags Becca as a high-risk candidate, and she is eliminated from the pool of potential applicants for a job.

On the vertical axis, we have secured real protections for Anna. However, neither the payment right nor the robust consent and inalienable or residual rights granted her do much of anything for Becca. And this is a problem, because Anna's act of uploading her fertility data puts her in a legally salient relationship with Becca. In a vital, fundamental sense, data about *Anna's* pregnancy is also

data about *Becca's* pregnancy; by which I mean that this data is used to make predictions and inform actions that directly impact Becca, too.

This example tells a story about data relations that are very specific: a person named Anna affecting a person named Becca. But in reality, data relations are not quite like this—they are population level relations arising from population-based correlations and patterns. So Anna is in a data relation not just with Becca, but potentially anyone that shares relevant, pregnancy-indicated features with her, and the same is true the other way around. These horizontal ties give rise to *population-level interests*. Insofar as either Becca or Anna have a legally salient interest in how pregnancy data is used to make decisions about them, that interest accrues at the population level: it attaches, or is implicated, by the choices of anyone who shares indicia of this population feature.

Data relations, like those between the population of Annas and the population of Beccas, are the point of data collection in the digital economy. Data relations are how money is made in the digital economy, why data is collected by companies like Apricot, and how that data is incorporated into the playbooks of success of the wealthiest companies in the world.

\* \* \* \*

Grounding our legal theories of data (and resulting debates in data governance law) in an accurate descriptive account of the data economy is helpful in several ways. The first is conceptual. It brings law and legal thinking into line with the underling activity it purports to regulate. If tech executives, data scientists, and others understand the value of data production as based on its ability to place people into population-based relations, then this is how lawyers and legal scholars should understand it, too.

The second is doctrinal. This relational account offers a strong challenge to the prevailing use of interpersonal contractual consent to govern data relations. To be sure, critiques of notice and consent-based privacy rules are far from new. But this account provides a particularly strong form of critique. Consent is not just a bad fit for data relations because these exchanges are particularly susceptible to manipulation, dark patterns, and information asymmetries, and contracts of adhesion—although all of these critiques have been persuasively and powerfully made about them. Even if we were to achieve perfect gold standard consent: applying it would to commit a category error.

I cannot consent on behalf of another, and they cannot consent on my behalf. But because of the pervasive and fundamental presence of data relations, that is what is happening all the time, constantly—constitutively—in the data economy. This suggests the need for other legal forms that capture the relevant legal interests at stake in data production, and that can meaningfully negotiate the terms of their legitimacy.

The third is institutional. Grasping the economic and theoretical significance of data relations opens up the possibility for a variety of legal mechanisms and legal institutions to grant recognition and standing to the legally relevant interests in data relations. Scholars and others proposing such institutions offer options that range in size and scope, as well as in form: they include data trusts, licensing schemes, imposing limits on current *de facto* monopoly rights in social data, enforcing tiered access rights to data, data destruction rules or agency auditing requirements. They may also include or coincide with innovations being proposed in the court system: changes to class actions and structural remedies, for instance.

The fourth is normative. A focus on data relations—and how law sets the terms that structure the quality of those relations—offers an alternative normative basis from which to evaluate the datafication of human life and when, or under what conditions, doing so is wrongful. On this account, datafication is not (only) wrong when it wrongfully renders the data subject legible, or denies the data subject her fair share of the profit her data exploitation generates. Instead, datafication is wrong if or when our data relations apprehend and entrench unjust social relations: by which I

mean data relations that enact or amplify forms of legally relevant social inequality. Thus, what might make Anna's data collection potentially wrongful is not (only) if Apricot collects it without her consent or uses it to manipulate Anna into buying something she does not need. Datafying Anna's pregnancy—or rather, the population of potentially pregnant people—is potentially wrongful if it rematerializes or amplifies the means by which pregnancy status is used to “do” pregnancy discrimination or contribute to systematic impoverishment and marginalization of women or other pregnant people. In other words, where data relations place people in positions of material or social subordination.

## REBECCA HAMILTON

Thank you Salomé. As you noted, the digital information economy vacuums up enormous quantities of data. How might we reconcile the potential conflict between data regulation that seeks to protect the privacy rights of individuals with the fact that violations of privacy go beyond the individual to the societal level? And thinking transnationally, what about the fact that societal level harms are likely to be viewed differently by different societies? If an individual approach to privacy protection is insufficient, what is the way forward? Asaf, with those questions in mind, let me turn to you next.

## REMARKS BY ASAF LUBIN<sup>2</sup>

<https://doi.org/10.1017/amp.2023.64>

I want to thank the family of Detlev Vagts for supporting the now annual Detlev F. Vagts Roundtable on Transnational Law and thank Rebecca for convening all of us. I also want to thank Salomé for producing such thought-provoking scholarship that indeed deserves our every praise, and I only hope my remarks will do it justice.

In opening my short statement, I thought it appropriate to revisit some of Professor Vagts' own work. In his 1979 co-authored *AJIL* article, “The Balance of Power in International Law: A History of an Idea,”<sup>3</sup> which Detlev actually wrote with his father Alfred, the two gentlemen claimed the following: “International Lawyers’ capacity to devise better and stronger institutions for keeping the peace will depend on their capacity to understand power structures and the unintended, but often inevitable, consequences that can flow from moves that exert pressure on one point in an interrelated system.”<sup>4</sup>

I want to propose a reading of Salomé's writing, in particular her groundbreaking *Yale Law Journal* article “A Relational Theory of Data Governance,”<sup>5</sup> which echoes the instructions left by Professor Vagts and his father Alfred. When Salomé writes about horizontal relations in the digital economy, what she is actually doing is uncovering the power structures that undergird our contemporary datasphere. She masterfully shows how datafication may lead to unfair wealth inequality that at once violates distributive ideals of justice and indexes, reproduces, and amplifies forms of social hierarchy.

<sup>2</sup> Dr. Asaf Lubin is an Associate Professor of Law at Indiana University Maurer School of Law and a Fellow at IU's Center for Applied Cybersecurity Research (CACR). He is additionally an affiliated fellow at Yale Law School's Information Society Project, a Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard University, and a visiting Scholar at the Hebrew University of Jerusalem Federmann Cyber Security Research Center.

<sup>3</sup> Alfred Vagts & Detlev F. Vagts, *The Balance of Power in International Law: A History of an Idea*, 73 *AJIL* 555 (1979).

<sup>4</sup> *Id.* at 580.

<sup>5</sup> Viljoen, *supra* note 1.