

On A Licensing Model For Consumer Data

Comments of Anant Raut¹

In this note, I propose a licensing framework for user data that more fairly rebalances the allocation of value received from the monetization of user data by an edge provider. The framework is modeled after music performance rights licensing in copyright law.

Where to draw a line in “big data”

When we talk about “big data” in tech, one of the issues that we debate is whether edge service providers should be allowed to collect as much user data and usage history as many do. Privacy-by-design proponents advocate collecting no more than the minimum amount of data necessary for the functioning of the edge service. The counterargument is that for users to continue to benefit from receiving many of these online services for free, edge providers have to make money somehow, which may involve collecting user data in order to sell targeted advertising on their sites, among other ways.

This counterargument, however, sweeps together the kind of data gathering that may be necessary to cover operational costs and data that is monetized for additional profit-making.

From the perspective of the edge provider, it is hard to draw a line between what profit-making is “necessary” and what is purely “surplus.” Innovation, expansion, security, and customer acquisition all require significant capital expenditure which must come from somewhere.²

There is no general agreement about whether a certain type of user data is inherently more valuable than others. There are existing laws that require greater protection for a subset of user data, personally identifiable information such as social security numbers; but some have argued that sheer volume of data can have a positive recursive effect allowing a search engine or ad placement program to become significantly better than its competitors, drawing more users and more advertisers, until it creates an insurmountable competitive advantage.

Privacy scholar Helen Nissenbaum theorized a contextual framework for data privacy, in which a privacy violation would be deemed to occur once the user data was shared in a different context than the one in which the data was first provided³, and that the user did not anticipate when providing his/her data initially.⁴ The problem with this framework *particular to privacy online* is that user expectations of privacy continue to evolve, downward. The Facebook newsfeed, hugely controversial when introduced for pushing out a rolling ticker of information about what

¹ Anant Raut is a former Counsel to two Assistant Attorneys General of the Department of Justice’s Antitrust Division. He also served as a special advisor to the Office of Vice President Joseph Biden and President Obama’s National Economic Council on stimulating market competition and real wage growth for low- and middle-income workers. He was previously a staff attorney in the Federal Trade Commission’s Bureau of Competition. These comments reflect his personal views.

² See my forthcoming Note on predatory considerations raised by venture capital-funded “free” services.

³ E.g., workplace versus health care

⁴ See, generally, Nissenbaum, H., *PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE*, Stanford University Press (2010).

one's contacts were up to, is now pedestrian. Online ads that suddenly appear hawking a product you searched for on an unrelated site have lost their element of shock, and are becoming accepted with resignation as an inescapable facet of an internet that seems to know who you are and where you are at all times. Defining the edges of a context in an administrable way in this setting may be impossible.

There is one bright line that can be easily drawn and is easy to administer, and that is between the edge service and third parties to whom it licenses, sells, or otherwise monetizes user data. It's pretty straightforward to draw a line between edge providers and third party entities.

There's also philosophical support within antitrust caselaw. Antitrust has traditionally looked favorably upon measures taken intra-firm to enhance competitiveness, and more skeptically with measures taken in conjunction with competing firms or outside parties.

Shifting from a first-sale to a licensing paradigm for consumer data

Underscoring the debate over how much user data is fair game for an edge service to collect is the sense among privacy advocates that users are being *exploited* by edge providers for their personal data. Their argument is that users are not receiving fair value for their data because edge providers are collecting vast troves of data on their users, which can be aggregated and monetized variously and repeatedly. In exchange, users often receive trivial or frivolous product.

Their argument is undergirded by edge providers that collect copious reams of data about their users while freely conceding that they have no specific plans on how they plan to use that data in the near future. This supports the idea that user data, individually and in aggregate, has significant value that can last well beyond the immediate transaction in which it is gathered.

If you accept the argument that users do not receive fair value for their data, then many of the policy issues raised by the collection of big data can be addressed by rebalancing the relationship between users and edge providers to compensate users more fairly for the value of their data. One solution would be to think of user data in a licensing framework, and require either the provision of additional value by the edge provider every time the edge provider sells, licenses, or otherwise monetizes the user's data to a third party, or affirmative opt-in.

How a licensing framework would work

The problem with current conceptions of consumer data is that we think of it the way we think of traditional mercantile exchanges, based upon first sale doctrine – that the user gives up his/her data, after which the edge provider is entitled to do whatever it wants with it. But the first sale approach is the wrong approach because it misses a fundamental aspect of user data, namely, that user data is not a tangible good, which can only be possessed by a single owner at a given time. Online data can be added, sorted, reconfigured, shared with multiple additional parties *simultaneously*, and replicated *infinitely and perfectly*. A better framework for reallocating value may be a licensing framework, such as exists in music licensing.

My proposed data licensing framework would require an edge service to provide additional value in exchange for the sale, licensing, or other monetization of user data to third parties, either in the form of compensation or additional service, OR obtain affirmative opt-in *every time*.

Under this approach, the edge provider with the direct user relationship may use the data internally without restriction to improve its product; in theory, this should be efficiency-enhancing, as it can help the edge provider make its service more useful to users, and become a stronger competitor. But when that data is shared, sold, or otherwise monetized with a third party, then each time the user should either receive additional compensation, OR affirmatively in.⁵ This compensation could take the form of fractions of fractions of pennies that accumulate over time. Or compensation could take the form of additional service and functionality from the direct edge provider.

(I added the alternative of an affirmative opt-in as a less onerous requirement than licensing, but it may not be necessary; the model could work purely as a licensing framework. The opt-in alternative gives edge providers a second way to monetize user data should they find the provision of additional value too difficult to administer.)

The system is a little onerous, but deliberately, to nudge edge providers into using user data to make improvements to their own product and user experience, and not to treat their users' data as windfall currency to be spent indiscriminately.

Analogues to the proposed framework exist in music copyright law

Providing additional value every time user data is monetized may seem nonviable and impossible to administer, especially having grown accustomed to a world in which there are practically no restrictions on how edge providers may use customer data. But such a framework already exists in music licensing. Every time a commercial establishment plays a song, it has to pay a performance royalty, which accumulates as fractions and fractions of a pennies, held by the artist's performance rights organization ("PRO") of choice, until enough has accumulated for the artist's PRO (ASCAP, BMI, or SESAC, in most circumstances) to send a check to the artist.

All of the arguments against a licensing framework for data can be made against a licensing framework for music and rebutted.

- *Aren't we talking about really small amounts of money?*

A: The long tail of performance rights royalties often involves tiny, fractional amounts of money that a PRO tallies until they reach an amount sufficiently large enough to merit sending a royalty check.

- *There are way too many users and edge providers to administer.*

A: The music performance rights royalty system is just as difficult to administer. There are hundreds of thousands of licensees, hundreds of thousands of licensors, and thousands upon thousands of songs being played at any given time around the country. Recent DOJ analysis of the consent decrees under which ASCAP and BMI have operated for decades

⁵ A stricter version of this model would require compensation under a licensing framework with no opt-in alternative. I leave a discussion of the comparative merits of the two approaches for a later time.

revealed a system in which the PROs could not even affirmatively say which performance rights were covered under their licenses. In spite of a vast number of administrative difficulties, the system still manages to compensate hundreds of thousands of singers for their songs.

- *Once the edge provider has the user data, it should be able to do whatever it wants with it.*

A: Once the licensor has the performance rights to a song, can that licensor turn around and resell the performance of that song infinitely to as many third parties as it wants? OF COURSE NOT, and this is the part that is most relevant in the context of consumer data. Each additional third party that wants to play the artist's song has to separately compensate the artist for his/her work.

Analogues to the proposed framework exist in labor law

Requiring some type of compensation in exchange for a thing of value from the user also has precedence in labor law, where non-compete agreements require additional consideration in order to be enforceable against an employee. Nor can continued employment (i.e., "sign this non-compete, or you no longer have a job") constitute the consideration; a number of jurisdictions prevent the party with superior bargaining power from making acceptance of all terms a binary choice. For purposes of this framework, an edge service would not be permitted to throw up a pop-up screen and require the user to agree to license his/her data in order to use the service at all. On the contrary, the European Union treats this type of behavior as a potential abuse of dominance, and is the basis of DG Comp's case against Facebook to unilaterally impose privacy requirements on its users. *Where the user has no meaningful choice, there can be no meaningful acceptance.*

Policy benefits of a licensing framework

From a policy perspective, the user gains more autonomy, is better informed, is better compensated, and has more control over the use of his/her data, *whichever* outcome the edge provider chooses. If the edge provider elects to compensate users for the monetization of their data under a licensing framework, then users become more aware of how their data is being used, and receive compensation. If the edge provider requests an affirmative opt-in, users have a sense of how frequently their data is being provided to third parties, and have a choice whether or not to provide it. If the edge provider decides against either alternative and only uses customer data internally, user data ends up being more secure, and users end up with a more useful and competitive product.

Addendum

One aspect of this that I reserve for a future discussion is how to determine compensation rates under the licensing framework. I will note generally that to the extent that the edge provider can find a way to price this data, the same information can be used to determine compensation for users based upon that price.