Facebook and Google Are the New Data Brokers

By Chris Hoofnagle | University of California, Berkeley

Facebook and Google Are the New Data Brokers

By Chris Hoofnagle | University of California, Berkeley

They claim not to sell data, but Facebook and Google have paid developers with data. This selling of your personal data is central to platform economics. Could quality signals change the terms of the bargain?

> "We do not sell your personal information to anyone"
> Google, Ads and Data Webpage

> "No, we don't sell any of your information to anyone and we never will."
> Facebook, Does Facebook sell my information? Webpage

Google and Facebook's claims about data selling are as false as they are adamant. To understand why one needs to change focus from the platform-advertiser relationship, and concentrate on developer-platform incentives. To grow their platforms, Google and Facebook reward developers—in effect, pay them—with personal information. But even experts in the field miss the point because information practices are opaque and misleading, and because most experience platforms as consumers rather than as developers.

This essay provides a structural critique focusing on the developer-platform relationship. The economic incentives of platforms make clear how and why platforms sell data to developers. It concludes by suggesting that light-touch quality signaling could help consumers understand how data are sold.

Developers and the Meaning of "Sell"

Teaching at Berkeley, I show students how developers get more and better data about users than one can obtain by simply using the web. Students learn in class that by using APIs, developers can investigate users and reveal interesting facts. We also show students how extensions and apps provide more data about users than when the same services are accessed through a browser. In fact, many apps add little value over a browser-intermediated experience. Yet, developers make apps that are simply skins of web browsers in order to get more data about a user than is available through the web and to enhance lock-in. The last semester I taught my privacy lab, students were most excited by Lumen, a Berkeley-ICSI developed tool that exposed how apps relay data back to app developers.

Developers' privileged access to user data is a "sale." While no money trades hands, a "sale" only requires a transfer of value. Sales include barters and promises to pay (e.g. using a check or credit card is still a sale, yet no money changes hands). To conceive of barter as not a sale would eviscerate the meaning of many laws including a core assumption of contract law. For instance, too clever by half criminals might "sell" a massage but bestow a sex act, or avoid "selling" drugs by trading them for guns. Such trades for value don't confuse the police, nor should they fool us.

Why Platforms Need Developers

Programmer development time and expertise is the most valuable resource at any technology firm. If programming were simple, emerging technology companies could outsource or even automate development and save on the lavish perks currently offered to employees. Outside developers are valuable force multipliers for platforms—in fact, outside developers are part of the definition of platform status.

Developers envision new functionalities and drive un-imagined forms of user growth and deepening "engagement" on platforms. This is why Cambridge Analytica developer Aleksandr Kogan was able to access so much personal information on Facebook. In exchange for getting just a few hundred thousand people to spend more time on Facebook completing surveys, Facebook was willing to give Kogan access to its "Open Graph," that is, your data and the data of your friends—tens of millions of them.

Now consider a different scenario: Kogan simply wanted to buy the same information in a data market. Data brokers will sell similar information, but the lower-bound cost for Kogan would be in thousands of dollars. That comparative cost gives us an idea of the "price" Facebook paid Kogan for user engagement.

One critical way to understand Google and Facebook is that both companies have a foot in the grave. Google's main innovation, PageRank, made the company an enormous fortune, but it is now primarily an advertising company, with multiple, increasingly desperate, failed attempts to diversify. We quickly forget all of Google's failed ventures and overlook that Google's big successes were acquisitions. Facebook became uncool some time ago, but a series of acquisitions such as Instagram, WhatsApp, and TBH ensured that the dissatisfied have nowhere to turn. In both cases these companies can use their enormous cash reserves to buy competitors, but as platforms they also have the ability to use developers to make the next killer app and keep the platform relevant.

Developers are so important that they go largely unmonitored and unpoliced by platforms. Early Facebook employee Katherine Losse explained in The Boy Kings that Zuckerberg treated developers as

trusted insiders, courting them with parties and "look[ing] away from the fact that almost all of Facebook users' data was available to them through the platform."

Google has similar incentives took away from developer problems. Recently the *Wall Street Journal* reported that Google was unable to even determine what happened to platform data exposed to developers, because it lacked audit rights and because it "didn't call or visit with any of the developers." Public revelations of platforms' standard business practices are now creating a correction, with both Facebook and Google now clamping down on developer access.

Lack of developer oversight also gives companies plausible deniability when it comes to giving consumers notice of data misuse. The Google incident remains an "exposure" instead of a "breach" because of the special trust given to developers—a naïve one extended to personal data but never to the data that Google deems important, such as its intellectual property. In considering giving notice, a Google spokesperson told the Journal that the company deliberated over "whether we could accurately identify the users to inform, whether there was any evidence of misuse, and whether there were any actions a developer or user could take in response…None of these thresholds were met here." Triggering these standards would be possible if Google were a data steward, instead of a nicely disguised data broker. Turning to intellectual property, note how deeply detailed Google's forensics were of the autonomous driving knowledge allegedly stolen by its employee who defected to Uber. In that context, Google can identify specific computers used, who used them, and the data taken; in the consumer data context there are advantages to not knowing: security events remain mere "exposures" rather than "breaches."

To mask the third-party nature of developers while giving them abundant access to data, Facebook and Google use misleading terms to describe these relationships. Facebook sometimes uses the term "service provider," and both Google and Facebook use "partner" in expansive, misleading ways that are incongruent with ordinary understandings of these terms. In Google parlance, "partner" means an arms-length, third-party relationship, but law requires partners to have a much closer, shared-risk/reward relationship. Many of us refer to our spouses as "partners," but under Facebook's use of the term, one could use the term for the dog walker.

When users complain about developer guile, platform companies respond with precise legal reasoning to pawn off blame, followed by naïve blubbering to cover up failure to anticipate and prevent foreseeable, obvious opportunism among developers. Consider how Facebook lawyer Paul Grewal explained away the Cambridge Analytica problem. Grewal placed the blame on users who downloaded the app, and explained that Facebook's policies were broken, that Kogan "lied to us." If developers were Facebook's partners in a legal sense, Facebook would have real remedies and leverage over Kogan. Grewal understands this because he understands words. He was, after all, the federal judge who presided over

Facebook's privacy cases before he joined Facebook as a VP. Business lawyers are supposed to anticipate and foreclose opportunism from contracting parties. Facebook's failure to do this is either one of the largest incidents of poor lawyering in the Valley, or calculated service in the effort to alienate data.

The platform-developer incentive structure also explains why, as Ezrachi and Stucke detail, Google bounced the privacy-preserving Disconnect and Professor Helen Nissenbaum's ad-flummoxing AdNauseum from its app store while allowing data-sucking apps such as "Brightest Flashlight" to remain. Disconnect and AdNauseum interfered with the data-selling transactions critical to Google's model; Brightest Flashlight perfectly exemplified what Google is offering to developers.

How Platforms Pay Developers with Your Data

How do the economic incentives work out in practice? In June 2018, the *New York Times* reported on a nice example of data trade for development dynamic. Facebook wanted to be on Blackberry devices. Blackberry wanted Facebook on its devices. But it would cost a small fortune and divert developer resources for Facebook to hire Blackberry OS programmers to build an app. Thus, Facebook made a deal: if BlackBerry would deeply integrate Facebook into its handsets, Facebook would make BlackBerry a "service provider" with privileged access to data, access so deep that it overrode user privacy settings.

How much was a freely-developed BlackBerry app worth to Facebook? And how much was liberal access to profiles, friend profiles, and Facebook functionality worth to BlackBerry? If you can understand those economic incentives, one can begin to understand how personal data is the currency platforms can offer. Facebook sold your data in order to get on Blackberry quickly and in order to avoid hiring a team of soon-to-be irrelevant BlackBerry OS experts.

Other examples abound. Until recently, if a company developed even a trivial webmail enhancement, Google would let them suck down inboxes. Developers who use "Sign in with Google" are rewarded with an API that lets them suck down the Google+ profile (name, email, and photo) of site visitors—Google even gives developers code so they don't have to figure out how to do it. Recall that for years, Google required users to sign up for Google+. No one wanted to. But Google wanted users to because it set up an environment where it could identify users by name everywhere online, and pawn the blame off on users for agreeing to create the profile.

Developers understand how they are paid with personal information. In fact, at least two Facebook developers (Six4three and Styleform) are suing the company for engaging in a bait and switch with user data. These developers made Facebook apps because the platform rewarded them with friend lists and pictures. Facebook began to restrict this access, thus destroying the economic model of these app

developers. But Facebook's restrictions did not end data sale; instead Facebook just became more strategic. Facebook created private APIs—like the Blackberry one—for special developers. Facebook's mobile first campaign was driven by access to consumer data. Accessing a service through a browser gives the user privacy advantages, as browsers limit information collection and tracking. Apps, on the other hand, leak voluminous amounts of data, so much so that the Mobile Marketing Association declared that "there is no anonymity" in mobile. Many mobile apps are simply skins—in essence the web site rendered inside an app. But that same experience enables the developer to obtain unique device identifiers and of course to create greater lock in.

Developer Access Matters

Does the transfer of this data to developers really matter? You might respond, "after all, it's only a machine seeing the data." But that response would never come out of the mouth of a programmer. Data scientists always look at raw data, in formats that are impossible to audit or secure.

Those who dismiss this access are also out of step with developments in ML, which show that by just looking at pictures, computers can recognize and infer secrets about individuals. We may soon prefer people to look at data as machines spend their evenings figuring out whether our children are gay or suicidal from videos we uploaded years ago. In fact, there will be no way to stop or even detect if Facebook or Google decide to train their computers to deeply analyze your children's physical features, voice, or even how they move their eyes as they interact with a device, and in so doing, reveal subtle hints about morbidity.

The whole "it's only a machine" debate is irrelevant anyway, because developers have security breaches all the time. Hundreds of times a month, data are dumped on the dark and deep web for humans and computers alike—for proof, just sign up for HaveIBeenPwned.

Some respond, "well, developers have strict contracts with platforms." But, these contracts are unenforceable because Google and Facebook can't even tell when they are violated. When they are violated, platforms engage in what we lawyers call "cover your ass," such as asking Cambridge Analytica to "certify" that it deleted data. Asking people who have violated trust to engage in a trust-based accountability mechanism is Facebook's only practical option because developers are not partners in any legal sense.

Some, like Facebook, respond with, "well, users gave permission." But platforms have a grotesque interpretation of consumer consent. To understand why, we need a structural and transaction cost analysis of the platform-user relationship. Yes, users give permission for data access, but the scope of

access, how permission is asked, the purposes for which data are used, the duration of the permission, and revocation of permission matter. Google and Facebook structure the transaction costs to encourage disclosure, and impose so many decision opportunities that privacy management is literally impossible.

Early in Facebook's history, developers were required to delete most user data within 24 hours. The 24 hour rule was an important check on "permission." In 2010, Facebook changed the policy, allowing developers to keep the data. Of course the 24 hour rule was widely ignored, and it created inefficiencies for developers. Yet, jettisoning the deletion rule rewarded developers with the creation of richer profiles, and denied users specific benefits. For instance, in a 24-hour retention world, a person wishing to leave Facebook could poison their profile, knowing that fake data would replace developers' profiles quickly. Under a more generous retention period, poisoning is harder - or impossible - to fully accomplish.
The poor quality of consent is illustrated by a study performed by Wijesekera et al., where researchers instrumented 36 android phones and gave them to users to use for a week. They found that Android phones pull data over 100,000 times a day per person, many of which occurred while the phone screen was blank. Google and Facebook made 6,000 location pulls a day. There is no doubt that users give permission for specific functionality, such as mapping. But Google controls the rules of access, and its priorities quickly make permission an illusion as it finds endless justifications to track you constantly.

Competitors can do better. A recent study found that Apple iPhones "communicated with Google domains at more than an order of magnitude (~50x) lower frequency than the Android device, and that Google did not collect any user location during the 24-hour experiment timeframe via iPhone." Furthermore, the iPhone itself communicates less often and with lower volume with Apple itself. Apple has a fundamentally different platform model, one that costs users more in up-front handset prices but then does much more to preserve user privacy. Apple also proves that one can have a top quality, information-enriched life without total information awareness.

The Need for a Quality Revolution

What can we do about platforms' incentives and data selling? To some extent, the problem is beginning to self-correct. Facebook realized some time ago that it needed to start limiting data selling and became more strategic about it. Now, reporting in the *Times* and *Journal* have brought all these relationships under pressure. But we are far from a virtuous cycle where platforms' incentives align with consumer interest.

More truthful signaling could help consumers choose alternatives more consistent with their preferences or at least make an implicit bargain explicit. Currently, platforms use language opportunistically and not consistently with legal or even popular understanding of terms. Policing the term "partner" alone would

deny platforms the ability to characterize third parties as something else. The drafters of the European General Data Protection Regulation (GDPR) understood this point and the Regulation forecloses many forms of definitional guile.

But more importantly, we need an honest definition of "sell" for the information industry. Sales of data include many kinds of transactions and if understood properly, include any situation where "consideration," that is value, is exchanged. This is the definition used the recently enacted California Consumer Privacy Act. It is not a radical, new definition. This definition of sale has long been used in contract law.

Now imagine how these simple, light-touch public policy interventions might trigger quality signals in these markets. If Android were meat, it would be stacked with the fatback and ham hocks, labeled grade B, close to its expiry date with viruses and pathogens bound to blossom. Facebook is even worse. Facebook would buy all the prime-grade competition and grind it into chuck. We need similar, simple signaling to make the quality bargain clear.

Data business models are inscrutable to most consumers, and it is easy to mislead the public about data practices because they are hidden, often protected by non-disclosure and even non-disparagement agreements to stop former employees from whistleblowing. Information-intensive companies can look you in the eye and tell you something truthful that is misleading.

It's time to wise up to one of the most misleading claims of Facebook and Google. Contrary to what they say, Facebook and Google sell your data. They don't sell it to advertisers. They sell it to developers. Here I have described the economic incentives and mechanisms of this data sale.

Facebook and Google are the modern data brokers. But they are worse than Choicepoint and Acxiom, because traditional data brokers never presumed that you agreed to their practices. Platforms blame you for being tricked about data sharing, they use weak accountability mechanisms for this sharing, they use their platforms to ply you into agreeing to boundless data selling, and when called out for it, platforms use language manipulatively, even mendaciously. Their hidden activities make a privacy market impossible because quality cannot be signaled. It's time to make their implicit models explicit. Let's see their practices for what they are—a clever scheme to sell your data and to blame you for agreeing to it.