

THE HIDDEN PRICE OF DATA

Laura Veldkamp

Revealing data's true price can turn passive users into active suppliers who demand fair value

Data is the fuel for the artificial intelligence algorithms that have lifted stock markets to historic highs on the promise of transforming economies. But how do we determine data's value? Data is not mined from the ground, not forged in factories. It accumulates unseen as a by-product of modern life: Every search, click, or morning walk with a phone in your pocket leaves a residue of information that someone, somewhere, can use.

When a good has no observable price—like a government service, for example—we typically value it at cost. But data has no explicit cost. When a retailer logs sales or a map app notes your location, that is data production. Of course, firms spend lavishly to process, analyze, and transform data. They hire armies of data scientists and invest in computing infrastructure to extract patterns from the noise. But the underlying raw data is like exhaust fumes from our economic engine. How do we value something that just appears?

The truth is that data is not free. We are all paid data producers. Once we comprehend that data is produced through transactions, a deeper economic logic emerges. If a profit-maximizing firm values

the data it receives from customers, it has an incentive to encourage more transactions, because more transactions mean more data. Customers buy more when they pay less. Firms that fail to offer discounts will see customers turn to competitors that do. Thus profit-maximizing firms must *discount* their goods and services, not out of fairness, but to generate more sales and more data.

Most of the economy today operates under this hidden bargain. Every digital purchase, every app download, every click is a dual transaction: Consumers buy a good or service, and at the same time, they *sell* their data. The observable price—the amount of money that changes hands—is really the *net price* of these two exchanges. Firms get revenue and data; consumers get products and convenience.

Price bundling

Here's the problem: As customers, we do not know what price, what discount, we received for data. This makes it impossible to know whether we received enough. Consumers typically do not have the option to purchase a good without selling their data. Requiring two transactions at the same time—in this case, a data sale and a product purchase—is

what economists call bundling. By hiding the price of data, bundling ensures that customers get less.

Imagine arriving in a foreign country with a different currency. On arrival, you pay the equivalent of \$18 for a lunch that should cost \$3. After a few days, you learn when to haggle, when to walk away, and what price is fair. In the digital economy, we are perpetually that first-day tourist. We sell our data every time we browse or buy. But because the transaction is bundled, we never see the price. We can't learn from experience.

Regulations that require firms to *unbundle* transactions—to post both the price with the right to use the transaction data and the price for a private transaction—would throw light on the data market. Consumers could observe the data discount. Some might decide it's worth it; others might withhold their data unless the discount is substantial. Over time, consumers would evolve from naive tourists into savvy suppliers of data, demanding their share of the data economy gains.

The challenge for economists and policymakers is to turn data—an ambient, invisible asset—into something we can count, contain, and price. Economists have begun to build a data measurement tool kit. Each approach offers a different take on “value” and will be feasible in different situations.

Five approaches

First, the market prices approach. Some data is traded in open markets, on platforms such as Snowflake or Datarade, where data sets are bought and sold. But this data is not a representative sample of economically important data. Most firms will not sell you their most valuable data, because it's central to their competitive advantage. But for the subset of data represented in these marketplaces, the price is a tried-and-true signal of value.

Second, the revenue approach. This treats data like any other productive asset: worth whatever extra revenue it can generate. This method asks a counterfactual question: What would profits look like if a firm didn't have some data? This approach requires a model that can predict what profits would have been without data. In finance, this is feasible because we know investors use data to buy more assets that will generate high returns. In other settings, data may have multiple uses that are harder to measure and quantify.

Third, complementary inputs approach. One way to infer the value of a firm's data stock is to look at the resources it devotes to managing and exploiting that data. Data doesn't produce value on its own; it becomes productive only when paired with people and tools. If you know how much labor and computing power a firm engages to work with data, and how

much it costs, you can infer the implicit value of the data stock that makes the spending worthwhile. It's indirect, but the surest sign that something is valuable is that firms pay real money to use it.

Fourth, the correlated behavior approach. If data improves decisions, we should see that in the alignment between what people do and the reward for those choices. Economists can measure those correlations between actions and payoffs to estimate how much information decision-makers must have had. In consumer markets, that might mean tracking how accurately recommendations match purchases, or how accurately a firm stockpiles goods that will sell well. High covariance implies valuable data at work. This approach measures data by its behavioral footprint.

Finally, the cost-accounting approach. By instinct accountants just add up the bills to get data. To some extent the new United Nations System of National Accounts does this by counting purchased data sets as assets. The hitch is that most data isn't bought; it's bartered. Consumers “pay” with information when they buy goods or use digital services. Those implicit discounts rarely appear on the books. A true accounting of data's cost would have to impute the value of the dollars or cents knocked off each purchase to encourage more transactions and more data revelation.

It's the simplest approach in theory and the hardest in practice, because it asks us to see data transactions that have never been itemized. Unbundling data and goods transactions by requiring separate pricing for transactions with and without the rights to use the transactions' data would make cost accounting feasible.

Toward quantification

Together, these five approaches describe an invisible asset class. Each captures an aspect of data value: labor devoted, revenue earned, the precision of actions, a market price, or implicit cost. None is infallible, feasible in all cases, or holistic in its measurement. Measurement is always imperfect. However, to make informed choices and craft sound policy, we must move data from the realm of intuition to the realm of quantification. Until then, the economy runs on a resource whose price we can only guess at, and whose value Silicon Valley can freely exploit. **F&D**

LAURA VELDKAMP is the Leon G. Cooperman Professor of Finance & Economics at Columbia University's Graduate School of Business and author of *The Data Economy: Tools and Applications*.

“We must move data from the realm of intuition to the realm of quantification.”