



Why Data Should Be Shared



Viktor Mayer-Schönberger

Economic innovation in the AI age could stall without regulations that mandate access to data

Google is remarkably good at guessing from users' misspelled queries what they intended to type into its search engine. This is because it doesn't guess: The internet giant trained its spellchecker some 20 years ago with typos billions of users made when searching. No competitor could come close because no one else had access to a similar fire hose of relevant data. Today Google accounts for 9 in every 10 internet searches—and faces new restraints after a recent antitrust ruling.

Using data to innovate, as Google did, is known as the “feedback effect.” Big Tech companies benefit from it most because they have access to the most data: They can crunch it at their data centers, transform it into insights, and use these to improve their products and services.

AI is turbocharging the feedback effect and widening the imbalance between the data haves and have-nots. It takes massive amounts of information and processing power to train and tune AI models, which large internet platforms have in spades. And what they don't have, they can buy with the avalanche of capital seeking to invest in AI.

The consequences for dominance are clear. Six Big Tech companies—Alphabet (Google), Netflix, Meta, Apple,

Amazon, and Microsoft—account for almost half the world's internet traffic. Four of these—Alphabet, Microsoft, Meta, and Amazon—dominate AI computing capacity.

As more data lead to better products and services, the largest players attract more customers, generating even more data. The feedback effect leads to self-reinforcing market concentration dynamics that competitors less flush with data can't join.

Concentration effects

Economists have long worried about the effects of concentration. Economies of scale and scope suggest that larger firms produce more cheaply than their smaller competitors, increasing sales while dictating price and pocketing profits, as Joseph Schumpeter argued in 1942. Innovation is the best antidote for concentration: Better ideas lead to improved or even completely novel products. This is crucial to economic dynamism.

Yet it is increasingly difficult for conventional companies to challenge the data economy's dominant players. They often lack the processing power and technical skills, but most important is the lack of a data mindset: the realization that using data creates value. Many conventional companies collect data but underutilize it; surveys show that at least 80 percent of what's collected worldwide is never used. Companies that collect data but don't know how to use it see the value of their digital resources seep away. Their capacity to innovate suffers, and they fall farther behind more data-savvy companies.

Innovation not only stalls within conventional companies. Dominant data platforms eventually suffer, too. Economists such as the University of Chicago's Ufuk Akcigit have shown that companies often lose interest in innovation once they become dominant and instead prioritize protecting their market share. Without robust competition they no longer need to innovate to stay ahead. Instead, they can weaken their offerings and still keep their substantial market share, as the writer Cory Doctorow argues.

The threat of data concentration and

loss of economic dynamism is serious enough to warrant policies that prevent or at least mitigate this misfortune. But identifying the best policy intervention is tricky.

Competition regulation

Using antitrust and competition regulation to break up large data platforms tackles the symptoms but not the cause of data concentration. If authorities were to break up Meta, for instance, another large platform would likely take its place. Because it doesn't change the underlying dynamic that rewards those that can access and use the most data.

Similarly, policies that give individuals more control over their data—think of the EU's General Data Protection Regulation—routinely fail to counter data concentration. Surveys show that many people care about their personal data, but few exercise their right to control it. This points to a problem of collective action: People must spend time curating access to their data, but they get only limited benefits in return, even though there are collective benefits. Everyone waits for others to act, and nothing happens: Powerful platforms continue to use data at will.

Policies that grant legal ownership or some similar exclusion right over data face similar practical hurdles. And given the complexity of licensing, these policies may result in less data being accessible overall, negatively impacting innovation. Moreover, transaction costs are not spread evenly: Complex negotiations over use licenses disproportionately burden individuals and small start-ups, tilting the playing field even further toward large platforms.

But regulations mandating access to data, especially nonpersonal data, offer more promise. If cleverly designed, they can reduce transaction costs—no licenses need be negotiated—helping smaller firms gain access to data. If data holders can extract value only by using the data themselves, it motivates those who have it to use it, pushing more conventional firms to become data savvy. Such regulations advance data use—what's currently lacking—rather than data collection.

“Using antitrust and competition regulation to break up large data platforms tackles the symptoms but not the cause of data concentration.”

Ideation and innovation

Innovation benefits as well. Multiple players can apply their ideas to data, so that ideation, not data hoarding, is rewarded. Data access mandates also adhere more closely to economic principles of value generation: The secret is often in data's clever application, or use, rather than its collection. To employ an industrial-age metaphor, access mandates facilitate value extraction rather than possession of raw materials.

Data access mandates may sound novel, but they are not, as Thomas Ramge and I show in our 2022 book, *Access Rules*. Governments worldwide are already legally obligated to offer access to troves of data. The best example is the locational data made available by the GPS system, operated by the US military, and the EU's Galileo system. The availability of costless yet accurate positional data has not only improved safety for airplanes, ships, and cars, it has yielded more efficient and sustainable logistics. It begot a multibillion-dollar industry.

Laws in many jurisdictions require companies to make certain data public, from financial results to emissions data. In the EU, large digital platforms must now share some data with smaller competitors. In the US, meanwhile, antitrust settlements have repeatedly required companies to let competitors

access their data. Google had to do so just recently as part of an antitrust trial. But the most spectacular (and often overlooked) case stems from an antitrust settlement in the 1950s that required AT&T to let US firms use its transistor patents for free. Start-ups seized this opportunity, designing and crafting integrated circuits—essentially bootstrapping Silicon Valley and the digital age.

More generally, the very mechanism at the core of the patent system in most countries is based on free information access: Patent holders get exclusive use of their invention only for a limited time, and only if they share the details of their invention so that others can learn from it.

The value data can generate by fueling innovation will only increase as the world transitions toward a comprehensive data economy. Unfortunately, this will also strengthen concentration dynamics that have spillover costs for the economy at large. Many policy interventions have been suggested. Data access mandates hold the most promise. **F&D**

VIKTOR MAYER-SCHÖNBERGER

is Professor of Internet Governance and Regulation at Oxford University's Internet Institute and coauthor of Big Data: A Revolution that Will Transform How We Work, Live, and Think.