## Data: What It Is, What It Isn't, and How Misunderstanding It is Fracturing the Internet

Nick Clegg



When Ukraine's President Volodymyr Zelenskyy made his impassioned plea to the Russian people on the eve of Vladimir Putin's invasion of his country, he knew it wouldn't reach them through Russia's state-controlled TV news programs. But it might reach them through social media.

In early 2022, there were 80 million active Instagram accounts in Russia, and more than 20 million people accessing Facebook. Millions more used TikTok, Twitter and all manner of other free-to-use apps and services. At the same time, millions of Ukrainians were using social media to tell the world beyond their borders what was happening.

Within days, Russia was blocking or restricting access to Facebook, Instagram and Twitter, attempting to cut Russian citizens off from the interconnected world of the open internet. TikTok restricted its own services in response to Russia's newly imposed "fake news" law.

Social media helps ideas, news and experiences cross national and cultural borders. At the start of the war, nearly three-quarters of Facebook and Instagram users in Russia had at least one friend outside the country. More than 60% of Facebook users in Ukraine had a Facebook friend in Russia, and 90% had at least one friend in the rest of the world. This free flow of information — people sharing their experiences and opinions, and seeing events unfold from the point of views of those outside their borders — was the antithesis of Russia's state-controlled propaganda machine.

The Russia-Ukraine war will have seismic repercussions for many years to come. It has united the European Union and the West — and rejuvenated the NATO military alliance — at a time when the demands of domestic politics seemed to be pulling in the opposite direction. But it has also exploded the

idea that economic interdependence inevitably leads to peace and stability. It has cast further doubt about the onward march of globalization, after the huge repercussions of the 2008 financial crisis and many of the political shockwaves of the last decade. The threat now is of rapidly escalating "deglobalization," protectionism and nationalism.

This is playing out in the digital sphere too. The rise of an authoritarian internet model — with citizens segregated from the rest of the global internet and subject to extensive surveillance — presents a real risk to the open, accessible internet as we know it. This is how the internet works in China, and other countries have made similar moves to build digital walls at their national boundaries. Russia was already moving in this direction before the internet clampdown that accompanied its invasion of Ukraine.

As lawmakers across the globe devise a much-needed new generation of internet regulations, a worrying strain of digital nationalism has crept into the debate. Talk of "digital sovereignty" and "data localization"— asserting a nation's right to stop or limit the free flow of cross border data — is now commonplace. And these ideas increasingly underpin new laws. As they do, they chip away at the foundations of the open internet, which relies on the flow of data across borders.

And they do so on a false premise. Underpinning digital nationalism is a misunderstanding of what data is and how it creates value.

Public discourse about data often relies on mistaken assumptions and metaphors from the industrial era that shape the way the debate is framed. One of the most dangerous misconceptions about data — in policymaking terms, at least — is that data is the "new oil": a scarce resource to be hoarded, enriching those who own the most. But data is not a finite commodity to be owned and traded, pumped from the ground and burned in cars and factories. It is something else entirely. And its ability to circulate and flow across borders is fundamental to how it creates value.

So what is data? And why is it so valuable?



## Data isn't oil

Data and information are not quite the same thing. Data can be ordered, systematically interrogated and used for purposes that previously wouldn't have been possible. A history book about Medieval England will be chock full of fascinating information, but it often won't be quantifiable data. On the other hand, the Domesday Book, which surveyed and valued landed property in late 11th century England, was the most comprehensive exercise in data collection of its era — even if it wasn't a gripping page turner.

This distinction is particularly true of quantitative data, which has been fundamental to the development of administrative systems and the bureaucracy of the modern state. Modern epidemiology, for example, relies on the work of proto data pioneers such as William Farr, who developed the first national vital statistics system in the UK of the mid-1800s. Today's data scientists and researchers use large-scale data, artificial intelligence and machine learning applications to <u>study the brain, accelerate the production</u> <u>of Covid-19 vaccines</u>, and much more.

Data has both a commercial and societal benefit. But it is not as simple as "more data equals more value." Data is a non-rivalrous good, which is a technical way of saying it doesn't get used up when it's consumed. Burn oil and it's gone forever. Make use of a data point and it still exists to be used again. Having a great big stockpile of it is not in and of itself particularly valuable. It's what you do with it that counts. The value of data stems from the quality of insights it can produce. Think of it this way: a long list of random words has much less value than a beautifully-written poem. The same principle applies to databases — a large database that includes random data points has little to no value (certainly to organizations without highly sophisticated systems capable of analyzing it), while a small database with well-connected data points can have great value.

No value is derived from the mere collection or storage of data. Unlike oil, the value of data depends on the context within which it is placed. A database about people's clothing preferences is much more valuable to a clothing retailer than it is to a restaurant chain, and vice versa for a database full of people's dining preferences. Yet the source of the data in both cases could be the same individual with a profile in the Banana Republic and OpenTable websites.

And, unlike oil, the value of data reduces over time. That is, the value of this year's data is much greater than the value of last year's data, and so on. My telephone number from last year, which I have since changed, has almost zero value to advertisers or anyone else. If a company has a spreadsheet full of customers' phone numbers, at what point is it no longer valuable? Is it when 5% of the numbers are no longer correct? 10%? 15%? And even if the "value" is positive, how much value must exist in order to justify a business decision to invest in generating new data?

None of this is to deny for a moment that data is often extremely valuable for organizations who know how to make use of what they have — whether they are a Silicon Valley tech company, a German car manufacturer or a local fast food delivery service. Not to mention charities or government health ministries. Gathering, storing and analyzing data at scale is very valuable indeed when it is high quality data that you know how to gain relevant insights from. But its value is very different to that of a scarce natural resource, and it is not constrained by how often or where it is processed, utilized, and consumed.



## **The Interconnected Data Society**

While data has been gathered in one form or another for centuries, it is the onset of the internet age that fundamentally changed our ability to gather and utilize it, turbo-charging its impact on society. Vellum parchment gave us the Domesday Book, but copying data was a laborious manual task, so its impact was limited. But step by step, technology has allowed us to make more of it. From index card systems to IBM's automated machines, the invention of digital memory, microprocessors that could sort through that memory, the network, the internet, mobile technology and artificial intelligence — at each stage new technology has massively increased the utility of data.

More than 3.5 billion people use Meta's apps — Facebook, Instagram, WhatsApp and Messenger — every month. It's a staggering number, and it's worth dwelling on for a moment. It means that between a third and half of all human beings on Earth use them. And in doing so, they have access to an interconnected world of people, ideas, news, communities and commerce unconstrained by local or national boundaries. This scale of connectedness is unprecedented in human history.

It's hard to overstate how important digital services have become to today's global economy. Data-driven technologies have contributed to growth and improved living standards the world over. It isn't just that the internet provides a faster or more convenient way of connecting businesses to customers. Data-based products empower individuals, for example by allowing them to compare prices at the touch of a button. They dramatically reduce costs and inefficiencies for businesses. And datasets can be aggregated and cross-referenced to gather new insights and identify opportunities that would otherwise have been invisible.

Processing and analyzing large quantities of data is now fundamental to how organizations in every sector operate. It is commonly understood that data is associated with medicine, telecommunications, banking and transportation, not to mention the administering of public services. Likewise, every professional sport — from football to gymnastics to aquatics — relies today on detailed data analysis to evaluate performance, draft prospects, and advance training methods. Even Artificial Intelligence — the training of computer systems to do things that have traditionally required human intelligence to do — and its subset Machine Learning — where a computer system can learn and train itself without explicit programming — are no longer the domain of technology companies alone.

The proliferation of accessible data-driven tools has not only been a boon for the global economy overall, it has also helped to democratize access to it, leveling the playing field between small businesses and big corporations. With social media apps and user-friendly e-commerce websites, people can start businesses online without the need for a big bank loan to pay for major overheads like renting a shopfront or office space. And with personalized digital ads they can reach targeted audiences of potential customers for just a few dollars, rather than the deep pockets required for mass-market TV, radio or billboard campaigns. This is especially true for people in rural communities and developing economies, where people with enormous talent and potential have been held back by poor infrastructure and their remoteness from metropolitan economic centers.

This democratization of access has in large part been made possible because many of the digital tools people use to access the open internet — including Meta's apps — are free to use. And it is only possible for them to be made available for free because of business models based on data-driven advertising. If companies like Meta instead started charging users a fee for their core services, it would immediately exclude millions of people probably billions — from using them.

So-called "big data" — commonly understood to mean the combination of huge storage capacity with advanced processing power, often aided by machine learning systems — presents enormous opportunities, economically and socially. But like all technological revolutions, these opportunities are accompanied by a new range of risks, dilemmas and challenges. As with previous technological advances, societies need to agree the parameters in which technology can operate and put in place guardrails that enable the good and mitigate the bad. And governments and other institutions need to find ways to harness the opportunities big data presents in helping to address all manner of societal challenges.

Through its Data for Good program, established in 2017, Meta has been part of an unprecedented collaboration between technology companies, the public sector, universities, nonprofits and others to aid disaster relief, support economic recovery, and inform policy and decision making. In recent years, the Data for Good program has informed the delivery of <u>medical aid and financial relief in Ukraine</u>, vaccination initiatives in <u>the</u> <u>Caribbean</u>, Brazil and the Philippines, disaster response in Mozambique, and enabled pioneering studies into things like <u>economic connectedness</u>, <u>attitudes towards climate change</u>, and the <u>challenges facing small businesses</u> <u>globally</u>.



## The internet needs guardrails, not roadblocks

Governments and regulators across the world are now grappling with the wide range of issues thrown up by the swift onset of the digital age: from the bumper digital markets and services acts in the EU, to the antitrust debates in the United States, online harms legislation in the United Kingdom and Ireland, or data protection laws proposed in India and elsewhere. These debates are long overdue — many of the issues at stake are too important to be left to private companies alone, which is why Meta has been publicly advocating for regulation in a number of areas for some time now.

But we mustn't throw the baby out with the bathwater. For governments, demanding greater sovereignty over data is a natural and understandable impulse, especially as other nations do the same. Many are drawn to the idea that by establishing digital walls at their national borders they can prevent data generated by their citizens from supposedly being extracted by powerful interests based overseas — in effect, treating it like oil reserves and preventing it from being exported. But this idea ignores the aggregation and network benefits of cross-border data flows.

These data flows are fundamental to the way the internet operates. Fixating on where data is stored and processed is a red herring — its value can be derived regardless of where in the world it is stored.

Because of how the global internet was built and has evolved, international data transfers occur as part of almost every online communication or activity. The internet was built to be a decentralized patchwork of tens of thousands of different networks that connect with and "talk to" one another by using standard technical protocols. Each of these networks routes data around the globe. The networks are generally agnostic of the physical "journey" of the data and instead optimize routing in real time to reduce latency and increase network resilience. Data localization policies impose unnecessary costs and technical challenges on what should be efficiency-based decision-making processes, making them market blockers rather than the drivers of economic growth some imagine them to be.

The independent think tank Information Technology and Innovation Foundation (ITIF) found a direct link between restrictive data policies, lower economic productivity and increased prices. In <u>a report published last July</u>, it found:

Restricting data flows has a statistically significant impact on a nation's economy — sharply reducing its total volume of trade, lowering its

productivity, and increasing prices for downstream industries that increasingly rely on data.

Using a scale based on OECD market-regulation data, ITIF finds that a 1-point increase in a nation's data restrictiveness cuts its gross trade output 7 percent, slows its productivity 2.9 percent, and hikes downstream prices 1.5 percent over five years.

ITIF also estimated the impact of restrictive data policies introduced in specific countries:

For Indonesia, the model estimates that over the five years, its moresignificant data restrictions reduced GOVs [gross output volumes] by 7.8 percent, lowered productivity by 3.2 percent, and raised prices by 1.6 percent. In the case of Russia, its heightened data restrictions between 2013 and 2018 cost an estimated 4.9 percent reduction in trade volume, a 2.0 percent reduction in productivity, and a 1.0 percent increase in prices of goods and services on average nationally.

Data-based technologies create value in ways that are radically different from older forms of economic and social organization. The harnessing of data has transformed our society dramatically and has the potential to continue to do so long into the future. Many of the consequences will be wonderful, others will be damaging. The task at hand is to maximize the former and minimize the latter.

Yes, Meta has an obvious self-interest in this debate. It's a global company whose services depend on its ability to store, transfer, and process data at scale. But it is far from alone. Millions of businesses share data across borders — corporations and startups alike — and millions more rely on datadriven products. The products and services these companies provide have become a part of everyday life for billions of people. It is integral to every sector and every type of organization — from banking and travel, to government and scientific researchers. And it has enabled millions of small businesses to do things that were previously out of their reach — from trading internationally to accessing payroll services or project management tools.

Meta could disappear from the face of the Earth and there would still be an overwhelming argument against data localization policies because the open internet is a guarantee of prosperity and freedom of thought that should be preserved in perpetuity. The repercussions of fragmenting the global internet will be felt far beyond the primary-colored campuses of Silicon Valley and the glass and steel towers of multinational corporations.

The risk we face is that as digital nationalism reshapes the internet piece by piece — new digital border by new digital border — its fundamental character will change. With each new national restriction, the internet becomes a little less free, and the digital economy becomes a little bit more constrained. Slowly, the authoritarian internet replaces the open internet; and authoritarian values replace democratic ones online.

We need a counterweight. Democracies must recognize and actively promote and defend the idea of the open internet. The announcement earlier this year of an agreement to protect open data flows between the US and the EU is a necessary step, as are the principles enshrined in the <u>Declaration for the Future of the Internet</u> announced by the Biden administration and signed by more than 60 national governments earlier this year. Likewise, the <u>Copenhagen Pledge on Tech and Democracy</u> — which now has more than 90 signatories — is a strong commitment to make digital technologies work for, not against, democracy and human rights.

These initiatives are welcome signs of leadership from the democratic world. We need them to turn into concrete actions.

The World Trade Organization (WTO) could be a place to strike a deal between countries on data flows. In fact, leading WTO countries have taken steps towards forging a new global trade agreement on digital economy issues, including addressing the free flow of data and other provisions to facilitate cross-border electronic commerce. The 2019 <u>plurilateral joint</u> <u>statement on "e-commerce"</u> is the most significant of those steps, and has now been signed by 86 WTO countries that represent 90% of global trade.

The statement includes the US and China, but notably not India. As the world's largest democracy, India could play a pivotal role in the future of the open internet. Persuading India — and others — to join the e-commerce negotiations would be a significant breakthrough.

With so much technological progress taking place over a relatively short space of time, we undoubtedly need new rules of the road to govern the use of data in modern society. But as they are designed we need to be careful not to lose the benefits that today's data-driven technologies have created — or those that tomorrow's technologies could bring.

Data isn't oil, and that means that those who want to design systems to ensure it is properly and responsibly managed — whether private companies or government regulators — have to think about it differently. Data can be a force for good, and it needn't come at people's expense. It is absolutely possible to design proper technical architectures to maximize the benefits of the data economy while minimizing the possible harms.

We need to create guardrails for the internet, not roadblocks. The challenge for policymakers is crafting new rules that take advantage of the great benefits that data-driven technologies bring to their societies and economies, while keeping people safe and protecting their privacy. The more this can be done at a global level the better, so we avoid creating regulatory silos and barriers to the seamless flows of data that make the open internet possible.

The further fracturing of the global internet is not inevitable. But preventing it will take leadership from those who believe in the democratic values that have made the internet the great liberating tool that it is today.



Written by Nick Clegg President, Global Affairs at Meta. Former Deputy Prime Minister of the United Kingdom, leader of the Liberal Democrats, member of Parliament.