

Mirror worlds

A deluge of data is giving rise to a new economy. Ludwig Siegele asks how it will work

AN ARMY OF doppelgangers is invading the world. Digital copies of aircraft engines, wind turbines and other heavy equipment came first. Now the electronic ghosts of smaller and larger things are joining them in the virtual realm, from toothbrushes and traffic lights to entire shops and factories. Even humans have begun developing these alter egos. In America the National Football League is planning to design an electronic avatar for every player.

These “digital twins”, as geeks term them, are far more than replicas of the original. Think of them more as shadows that are, thanks to a multitude of sensors and wireless connectivity, intimately linked to their physical selves, and every day producing oceans of data. If something happens in the real world, it is rapidly reflected in this shadow realm. Some digital twins already come with the laws of nature programmed in. They double as a database of everything that has ever happened to the original. This makes it possible to look into their future. Sports coaches, for instance, will be able to run simulations, predict when an athlete might get injured and adjust training routines to avoid problems.

Digital twins are just one part of a vast shift in the world’s economy. They populate what David Gelernter of Yale University long ago forecast as “mirror worlds”: a new dimension of human life based on and fuelled by data. Year by year, ever more parts of the

physical realm are coming to be represented and simulated in the virtual world—an inversion of Plato’s theory that real-world objects are just imperfect copies of their true being in the spiritual realm. The emergence of these mirror worlds will bring about a distinct economy. This development will require new markets, institutions, infrastructure, businesses and even geopolitical arrangements. It is the promises and pitfalls of the new “data economy” which will be the focus of this special report.

Mirror worlds are not mere mathematical representations of real ones. They also give new meaning to the adage that knowledge is power. Increasingly, digital copies are taking on lives of their own and acting on the physical world. They can be used to optimise everything, from the acoustics of a headset to an entire national railway network. They will enable all sorts of artificial-intelligence (AI) algorithms to recognise objects and faces, understand speech and even distinguish smells. And they make possible new business models: why buy heavy equipment if its wear and tear can be measured in detail and it can thus be rented by the minute?

A good place to start analysing any economy is by measuring it. A robust methodology has yet to be developed, but the data economy is already large. Statistics Canada, a government agency, last year tried to estimate the value of the country’s data (its stock plus ►►

related software and intellectual property in the field). The result was between c\$157bn and c\$218bn (\$118bn and \$164bn). If that number is close—a big “if”—the value of all the data in America, whose GDP is 12 times that of Canada, could amount to \$1.4trn-2trn, which would be nearly 5% of America’s stock of private physical capital.

If the amount of data generated around the world is any guide, this new economy is growing fast. The first human genome (three gigabytes of data, which nearly fills a DVD) was sequenced 17 years ago; in April, 23andMe, a firm which offers genetic testing, claimed more than 10m customers. The latest autonomous vehicles produce up to 30 terabytes for every eight hours of driving (or some 6,400 DVDs). IDC, a market-research firm, estimates the world will generate about 90 zettabytes (19trn DVDs) this year and next (see chart), more than all data produced since the advent of computers.

Yet even more striking than the rapid growth of the data economy are the tensions and trade-offs it produces. Take its economics. In some ways, data are a natural resource, much like oil, which can be owned and traded (this newspaper called data the “world’s most valuable resource” in 2017). But data also have characteristics of a public good, which ought to be used as widely as possible to maximise wealth creation. New institutions must be created to reflect this tension, as was the case for intellectual property.

The infrastructure of the data economy, too, is torn between two poles. Currently, it mainly consists of huge data centres packed with servers where data are stored and crunched. Yet such centralisation has drawbacks, not least because it consumes huge amounts of energy and creates privacy risks. A decentralising counter-movement is already under way: more data are processed at the “edge”, closer to where they are collected.

Businesses are also facing a digital reversal. Many firms want to use data to infuse their corporate applications with AI. They have built central repositories such as “data lakes”, which hold all kinds of digital information. Such systems are of limited use, however, if a firm and its employees lack the required skills, refuse to believe the data or even to share them internally.

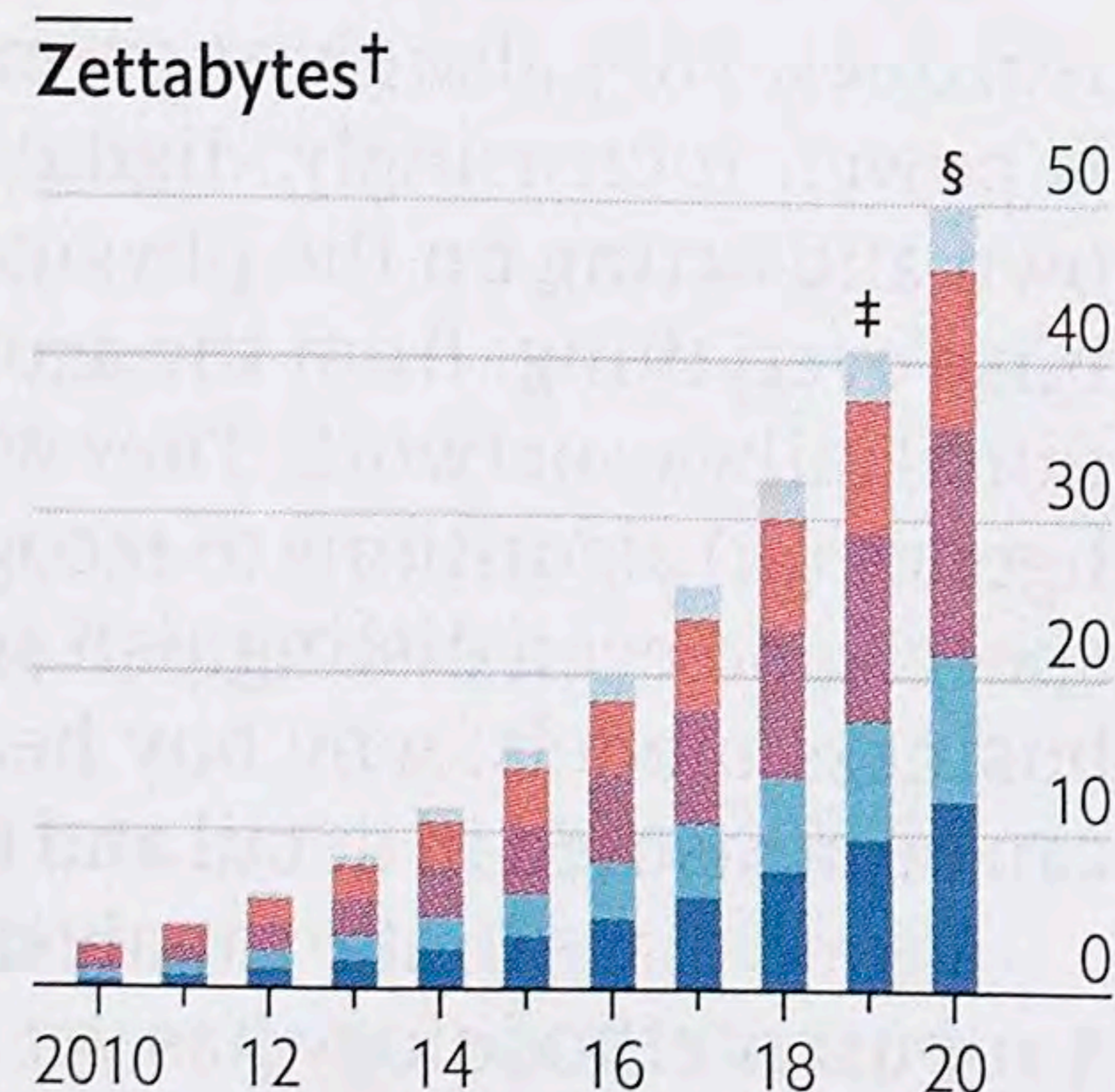
Finally, the geopolitics of data will not be simple, either. Online giants in particular have assumed that the data economy will be a global affair, with the digital stuff flowing to where processing is best done for technical and cost reasons. Yet governments are increasingly asserting their “digital sovereignty”, demanding that data not leave their country of origin.

This special report will tackle these topics in turn. It will conclude by discussing what is perhaps the biggest conundrum of the mirror world: the risk is that the wealth it creates will be even more unequally distributed than in its terrestrial twin. ■

Deluged

Data generated, worldwide

Average annual increase
2010-18, %



*Europe, Middle East and Africa

†1ZB=1 trillion GB ‡Estimate §Forecast

Source: IDC, Seagate

Economics

Digital plurality

Are data more like oil or sunlight?

PASSIONATE GRAMMARIANS have long quarrelled over whether data should be singular or plural (contrary to common usage, this newspaper is sticking with the latter, for now). A better question is why are data so singularly plural? That is, why do they have so many different faces?

For an answer, start with the many metaphors used to describe flows of data. Originally they were likened to oil, suggesting that data are the fuel of the future. More recently, the comparison has been with sunlight because soon, like solar rays, they will be everywhere and underlie everything. There is also talk of data as infrastructure: they should be seen as a kind of digital twin of roads or railways, requiring public investment and new institutions to manage them.

The multiplication of metaphors reflects the malleable economics of data. First, they are “non-rivalrous”: since they are infinitely copyable, they can be used by many people without limiting the use by others. But they are also “excludable”: technologies like encryption can control who has access to them. Depending on where one sets the cryptographic slider, data can indeed be private goods like oil or public goods like sunlight—or something in between, known as a “club good”.

This in turn means that there is not just one data economy, but three more or less distinct ones, each with its own ideology. And the big question is whether one will come to dominate, or whether the mirror world will be as much of a mixture as the real one.

If oil is still the most-used metaphor, it is because comparing data to the black stuff is easy. Like oil, data must be refined to be useful. In most cases they need to be “cleansed” and “tagged”, meaning stripped of inaccuracies and marked to identify what can be seen, say, on a video. This has spawned a global industry employing hundreds of thousands of people, mostly in low-wage countries. Scale AI, a startup in San Francisco, employs 30,000 taggers around the world who review footage from self-driving cars and ensure the firm’s software has correctly classified things like houses and pedestrians.

Before data can power AI services, they also need to be fed through algorithms, to teach them to recognise faces, steer self-driving cars and predict when jet engines need a check-up. And different data sets often need to be combined for statistical patterns to emerge. In the case of jet engines, for instance, mixing usage and weather data helps forecast wear and tear.

The oil metaphor also rings true because some types of data and some of the insights extracted from them are already widely traded. Online advertising is perhaps the biggest marketplace for personal data: clicks are bought and sold based on a detailed digital profile of each viewer. It was worth \$178bn globally in 2018, according to Strategy&, a consultancy. Data brokers, which can track thousands of data points for each individual, do brisk business with personal information, too. They sell it to everyone from banks to telecoms carriers, generating annual revenue of more than \$21bn, says Strategy&.

Offering insights from mining data can be very profitable, too. On Kaggle, a website owned by Google that hosts machine-learning contests, thousands of teams of data scientists compete against each other to see who can come up with the best algo- ▶▶

"Mirror worlds." *Economist*, February 22nd-28th 2020, pp. 3+. The Economist Historical Archive, link.gale.com/apps/doc/TPIVUD656099404/ECON?u=crepuq_mcgill&sid=bookmark-ECON. Accessed 1 Nov. 2023.