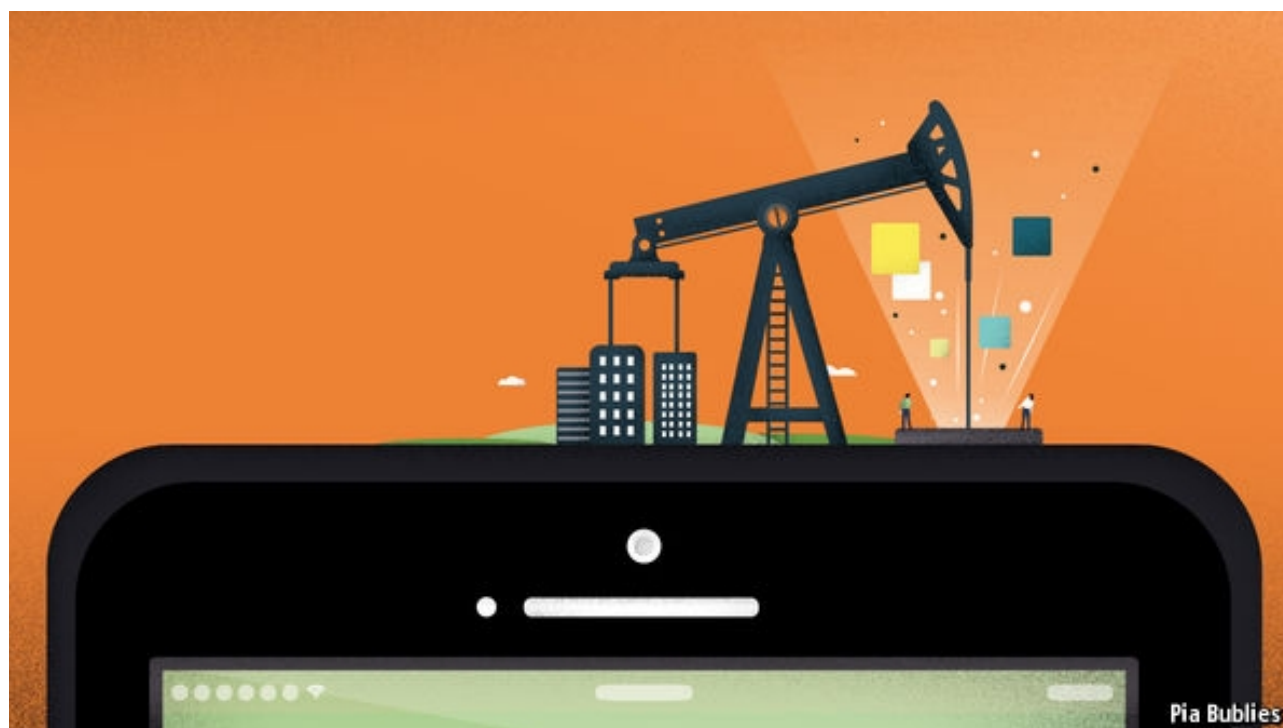


Fuel of the future

Data is giving rise to a new economy

How is it shaping up?



Print edition | Briefing

May 6th 2017

AN OIL refinery is an industrial cathedral, a place of power, drama and dark recesses: ornate cracking towers its gothic pinnacles, flaring gas its stained glass, the stench of hydrocarbons its heady incense. Data centres, in contrast, offer a less obvious spectacle: windowless grey buildings that boast no height or ornament, they seem to stretch to infinity.

Yet the two have much in common. For one thing, both are stuffed with pipes. In refineries these collect petrol, propane and other components of crude oil, which have been separated by heat. In big data centres they transport air to cool tens of

thousands of computers which extract value—patterns, predictions and other insights—from raw digital information.

Latest updates

A lesson in parkonomics, for Jeremy Corbyn
SPEAKERS' CORNER

Angela Merkel should seize this chance to remake Europe
KAFFEEKLATSCH

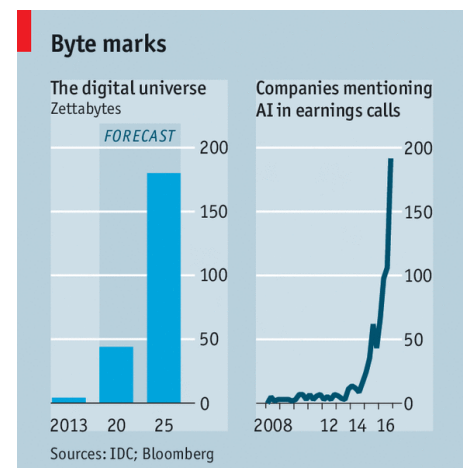
When flyers end up in the wrong cities
GULLIVER

See all updates

Both also fulfil the same role: producing crucial feedstocks for the world economy. Whether cars, plastics or many drugs—without the components of crude, much of modern life would not exist. The distillations of data centres, for their part, power all kinds of online services and, increasingly, the real world as devices become more and more connected.

Data are to this century what oil was to the last one: a driver of growth and change. Flows of data have created new infrastructure, new businesses, new monopolies, new politics and—crucially—new economics. Digital information is unlike any previous resource; it is extracted, refined, valued, bought and sold in different ways. It changes the rules for markets and it demands new approaches from regulators. Many a battle will be fought over who should own, and benefit from, data.

There is an awful lot to scrap over. IDC, a market-research firm, predicts that the “digital universe” (the data created and copied every year) will reach 180 zettabytes (180 followed by 21 zeros) in 2025 (see chart). Pumping it all through a broadband internet connection would take over 450m years. To speed the transfer into its data centres, Amazon, an e-commerce giant with a fast-growing cloud-computing arm, uses trucks pulling shipping containers each packed with storage devices holding 100 petabytes (a mere 15 zeros). To ingest it all, firms are speedily building data refineries. In 2016 Amazon, Alphabet and Microsoft together racked up nearly \$32bn in capital expenditure and capital leases, up by 22% from the previous year, according to the *Wall Street Journal*.



Economist.com

The quality of data has changed, too. They are no longer mainly stocks of digital information—databases of names and other well-defined personal data, such as age, sex and income. The new economy is more about analysing rapid real-time flows of often unstructured data: the streams of photos and videos generated by users of social networks, the reams of information produced by commuters on their way to work, the flood of data from hundreds of sensors in a jet engine.

From subway trains and wind turbines to toilet seats and toasters—all sorts of devices are becoming sources of data. The world will bristle with connected sensors, so that people will leave a digital trail wherever they go, even if they are not connected to the internet. As Paul Sonderegger, a big-data strategist at Oracle, a software-maker, puts it: “Data will be the ultimate externality: we will generate them whatever we do.”

It is what you know

Most important, the value of data is increasing. Facebook and Google initially used the data they collected from users to target advertising better. But in recent years they have discovered that data can be turned into any number of artificial-intelligence (AI) or “cognitive” services, some of which will generate new sources of revenue. These services include translation, visual recognition and assessing someone’s personality by sifting through their writings—all of which can be sold to other firms to use in their own products.

Although signs of the data economy are everywhere, its shape is only now becoming clear. And it would look pretty familiar to J.R. Ewing. There are the data majors, a growing number of wildcatters and plenty of other firms trying to get a piece of the action. All are out to exploit a powerful economic engine called the “data-network effect”—using data to attract more users, who then generate more data, which help to improve services, which attracts more users.

The majors pump from the most bountiful reservoirs. The more users write comments, “like” posts and otherwise engage with Facebook, for example, the more it learns about those users and the better targeted the ads on newsfeeds become. Similarly, the more people search on Google, the better its search results turn out.

These firms are always looking for new wells of information. Facebook gets its users to train some of its algorithms, for instance when they upload and tag pictures of friends. This explains why its computers can now recognise hundreds of millions of people with 98% accuracy. Google's digital butler, called "Assistant", gets better at performing tasks and answering questions the more it is used.

Uber, for its part, is best known for its cheap taxi rides. But if the firm is worth an estimated \$68bn, it is in part because it owns the biggest pool of data about supply (drivers) and demand (passengers) for personal transportation. Similarly, for most people Tesla is a maker of fancy electric cars. But its latest models collect mountains of data, which allow the firm to optimise its self-driving algorithms and then update the software accordingly. By the end of last year, the firm had gathered 1.3bn miles-worth of driving data—orders of magnitude more than Waymo, Alphabet's self-driving-car division.

"Data-driven" startups are the wildcatters of the new economy: they prospect for digital oil, extract it and turn it into clever new services, from analysing X-rays and CAT scans to determining where to spray herbicide on a field. Nexar, an Israeli startup, has devised a clever way to use drivers as data sources. Its app turns their smartphones into dashcams that tag footage of their travels via actions they normally perform. If many unexpectedly hit the brake at the same spot on the road, this signals a pothole or another obstacle. As compensation for using Nexar's app, drivers get a free dashcam and services, such as a detailed report if they have an accident. The firm's goal is to offer all sorts of services that help drivers avoid accidents—and for which they, or their insurers, will pay. One such is alerts about potholes or when a car around a blind corner suddenly stops.

Non-tech firms are trying to sink digital wells, too. GE, for instance, has developed an "operating system for the industrial internet", called Predix, to help customers control their machinery. Predix is also a data-collection system: it pools data from devices it is connected to, mixes these with other data, and then trains algorithms that can help improve the operations of a power plant, when to maintain a jet engine before it breaks down and the like.

As in oil markets, bigger data firms keep taking over smaller ones (see table). But another aspect of the data economy would look strange to dealers in black gold. Oil

is the world's most traded commodity by value. Data, by contrast, are hardly traded at all, at least not for money. That is a far cry from what many had in mind when they talked about data as a “new asset class”, as the World Economic Forum (WEF), the Davos conference-organiser-cum-think-tank, did in a report published in 2011. The data economy, that term suggests, will consist of thriving markets for bits and bytes. But as it stands, it is mostly a collection of independent silos.

Keep it to yourself

This absence of markets is the result of the same factors that have given rise to firms. All sorts of “transaction costs” on markets—searching for information, negotiating deals, enforcing contracts and so on—make it simpler and more efficient simply to bring these activities in-house. Likewise, it is often more profitable to generate and use data inside a company than to buy and sell them on an open market.

Target company (Year)	Value of deal, \$bn	Business
Instagram (2012)	1.0	Photo sharing
WhatsApp (2014)	22.0	Text/photo messaging
Waze (2013)	1.2	Mapping and navigation
The Weather Company (2015)	2.0	Meteorology
Truven Health Analytics (2016)	2.6	Health care
Mobileye (2017)	15.3	Self-driving cars
SwiftKey (2016)	0.25	Keyboard/artificial intelligence
LinkedIn (2016)	26.3	Business networking
BlueKai (2014)	0.4	Cloud data platform
DataLogix (2014)	1.0	Marketing

Source: Company reports, estimates
Economist.com

Their abundance notwithstanding, flows of data are not a commodity: each stream of information is different, in terms of timeliness, for example, or how complete it may be. This lack of “fungibility”, in economic lingo, makes it difficult for buyers to find a specific set of data and to put a price on it: the value of each sort is hard to compare with other data. There is a disincentive to trade as each side will worry that it is getting the short end of the stick.

Researchers have only just begun to develop pricing methodologies, something Gartner, a consultancy, calls “infonomics”. One of its pioneers, Jim Short of the University of California in San Diego, studies cases where a decision has been made about how much data are worth. One such involves a subsidiary of Caesars Entertainment, a gambling group, that filed for bankruptcy in 2015. Its most valuable asset, at \$1bn, was determined to be the data it is said to hold on the 45m customers who had joined the company's customer-loyalty programme over the previous 17 years.

The pricing difficulty is an important reason why one firm might find it simpler to buy another, even if it is mainly interested in data. This was the case in 2015 when IBM reportedly spent \$2bn on the Weather Company, to get its hands on mountains

of weather data as well as the infrastructure to collect them. Another fudge is barter deals: parts of Britain's National Health Service and DeepMind, Alphabet's AI division, have agreed to swap access to anonymous patient data for medical insights extracted from them.

The fact that digital information, unlike oil, is also “non-rivalrous”, meaning that it can be copied and used by more than one person (or algorithm) at a time, creates further complications. It means that data can easily be used for other purposes than those agreed. And it adds to the confusion about who owns data (in the case of an autonomous car, it could be the carmaker, the supplier of the sensors, the passenger and, in time, if self-driving cars become self-owning ones, the vehicle itself).

“Trading data is tedious,” says Alexander Linden of Gartner. As a result, data deals are often bilateral and ad hoc. They are not for the fainthearted: data contracts often run over dozens of pages of dense legalese, with language specifying allowed uses and how data are to be protected. A senior executive of a big bank recently told Mr Linden that he has better things to do than sign off on such documents—even if the data have great value.

In the case of personal data, things are even more tricky. “A regulated national information market could allow personal information to be bought and sold, conferring on the seller the right to determine how much information is divulged,” Kenneth Laudon of New York University wrote in an influential article entitled “Markets and Privacy” in 1996. More recently, the WEF proposed the concept of a data bank account. A person's data, it suggested, should “reside in an account where it would be controlled, managed, exchanged and accounted for”.

The idea seems elegant, but neither a market nor data accounts have materialised yet. The problem is the opposite to that with corporate data: people give personal data away too readily in return for “free” services. The terms of trade have become the norm almost by accident, says Glen Weyl, an economist at Microsoft Research. After the dotcom bubble burst in the early 2000s, firms badly needed a way to make money. Gathering data for targeted advertising was the quickest fix. Only recently have they realised that data could be turned into any number of AI services.

Slave to the algorithm

Whether this makes the trade of data for free services an unfair exchange largely depends on the source of the value of these services: the data or the algorithms that crunch them? Data, argues Hal Varian, Google's chief economist, exhibit "decreasing returns to scale", meaning that each additional piece of data is somewhat less valuable and at some point collecting more does not add anything. What matters more, he says, is the quality of the algorithms that crunch the data and the talent a firm has hired to develop them. Google's success "is about recipes, not ingredients."

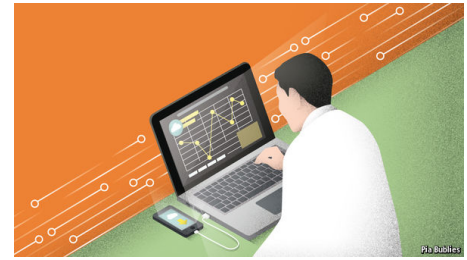
That may have been true in the early days of online search but seems wrong in the brave new world of AI. Algorithms are increasingly self-teaching—the more and the fresher data they are fed, the better. And marginal returns from data may actually go up as applications multiply, says Mr Weyl. After a ride-hailing firm has collected enough data to offer one service—real-time traffic information, say—more data may not add much value. But if it keeps collecting data, at some point it may be able to offer more services, such as route planning.

Such debates, as well as the lack of a thriving trade in data, may be teething problems. It took decades for well-functioning markets for oil to emerge. Ironically, it was Standard Oil, the monopoly created by John D. Rockefeller in the late-19th century, that speeded things up: it helped create the technology and—the firm's name was its programme—the standards that made it possible for the new resource to be traded.

Markets have long existed for personal data that are of high value or easy to standardise. So-called "data brokers" do a swift trade in certain types of data. In other areas, markets, or something akin to them, are starting to develop. Oracle, which dominates the market for corporate databases, for example, is developing what amounts to an exchange for data assets. It wants its customers to trade data, combine them with sets provided by Oracle and extract insights—all in the safe environment of the firm's computing cloud, where it can make sure, among other things, that information is not misused. Cognitive Logic, a startup, has come up with a similar product, but leaves the data in separate IT systems.

Other young firms hope to give consumers more of a stake in their data. Citizenme allows users to pull all their online information together in one place and earn a small fee if they share it with brands. Datacoup, another startup, is selling insights from personal data and passing on part of the proceeds to its users.

So far none of these efforts has really taken off; those focusing on personal data in particular may never do so. By now consumers and online giants are locked in an awkward embrace. People do not know how much their data are worth, nor do they really want to deal with the hassle of managing them, says Alessandro Acquisti of Carnegie Mellon University. But they are also showing symptoms of what is called “learned helplessness”: terms and conditions for services are often impenetrable and users have no choice than to accept them (smartphone apps quit immediately if one does not tap on “I agree”).



For their part, online firms have become dependent on the drug of free data: they have no interest in fundamentally changing the deal with their users. Paying for data and building expensive systems to track contributions would make data refiners much less profitable.

Data would not be the only important resource which is not widely traded; witness radio spectrum and water rights. But for data this is likely to create inefficiencies, argues Mr Weyl. If digital information lacks a price, valuable data may never be generated. And if data remain stuck in silos, much value may never get extracted. The big data refineries have no monopoly on innovation; other firms may be better placed to find ways to exploit information.

The dearth of data markets will also make it more difficult to solve knotty policy problems. Three stand out: antitrust, privacy and social equality. The most pressing one, arguably, is antitrust—as was the case with oil. In 1911 America’s Supreme Court upheld a lower-court ruling to break up Standard Oil, which then controlled around 90% of oil refining in the country.

Some are already calling for a similar break-up of the likes of Google, including Jonathan Taplin of the University of Southern California in his new book “Move

Fast and Break Things”. But such a radical remedy would not really solve the problem. A break-up would be highly disruptive and slow down innovation. It is likely that a Googlet or a Babyface would quickly become dominant again.

Yet calls for action are growing. The “super-platforms” wield too much power, says Ariel Ezrachi of the University of Oxford, who recently published a book entitled “Virtual Competition” with Maurice Stucke of the University of Tennessee. With many more and fresher data than others, he argues, they can quickly detect competitive threats. Their deep pockets allow them to buy startups that could one day become rivals. They can also manipulate the markets they host by, for example, having their algorithms quickly react so that competitors have no chance of gaining customers by lowering prices (see [Free exchange \(http://www.economist.com/news/finance-and-economics/21721648-trustbusters-might-have-fight-algorithms-algorithms-price-bots-can-collude\)](http://www.economist.com/news/finance-and-economics/21721648-trustbusters-might-have-fight-algorithms-algorithms-price-bots-can-collude)). “The invisible hand is becoming a digital one,” says Mr Ezrachi.

Beware the digital hand

At a minimum, trustbusters have to sharpen their tools for the digital age. The European Commission did not block the merger of Facebook and WhatsApp. It argued that although these were operating the two largest text-messaging services, there were plenty of others around and that the deal would also not add to Facebook’s data hoard because WhatsApp did not collect much information about its users. But Facebook was buying a firm that it feared might evolve into a serious rival. It had built an alternative “social graph”, the network of connections between friends, which is Facebook’s most valuable asset. During the approval process of the merger Facebook had pledged that it would not merge the two user-bases, but started doing so last year, which has led the commission to threaten it with fines.

The frustration with Facebook helps explain why some countries in Europe have already started to upgrade competition laws. In Germany legislation is winding through parliament which would allow the Federal Cartel Office to intervene in cases in which network effects and data assets play a role. The agency has already taken a special interest in the data economy. It has launched an investigation into whether Facebook is abusing its dominant position to impose certain privacy policies. Andreas Mundt, its president, wants to do more: “Can we further optimise

our investigation techniques? How can we better integrate dynamic effects into our analyses?”

A good general rule for regulators is to be as inventive as the companies they keep an eye on. In a recent paper Messrs Ezrachi and Stucke proposed that antitrust authorities should operate what they call “tacit collusion incubators”. To find out whether pricing algorithms manipulate markets or even collude, regulators should run simulations on their own computers.

Another idea is to promote alternatives to centralised piles of data. Governments could give away more of the data they collect, creating opportunities for smaller firms. They could also support “data co-operatives”. In Switzerland a project called Midata collects health data from patients, who can then decide whether they want them to be included in research projects.

Distributing the data

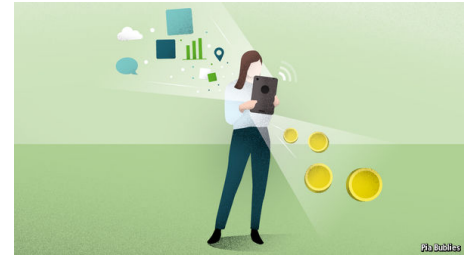
For some crucial classes of data, sharing may even need to be made mandatory. Ben Thompson, who publishes *Stratechery*, a newsletter, recently suggested that dominant social networks should be required to allow access to their social graphs. Instagram, a photo-sharing service which has also been swallowed by Facebook, got off the ground by having new users import the list of their followers from Twitter. “Social networks have long since made this impossible, making it that much more difficult for competitors to arise,” Mr Thompson points out.

Mandatory data sharing is not unheard of: Germany requires insurers jointly to maintain a set of statistics, including on car accidents, which smaller firms would not be able to compile on their own. The European Union’s new General Data Protection Regulation (GDPR), which will start to apply in May 2018, requires online services to make it easy for customers to transfer their information to other providers and even competitors.

But “data portability”, as well as data sharing, highlights the second policy problem: the tension between data markets and privacy. If personal data are traded or shared they are more likely to leak. To reduce this risk, the GDPR strengthens people’s control over their data: it requires that firms get explicit consent for how they use

data. Fines for violations will be steep: up to 4% of global revenues or €20m (\$22m).

Such rules will be hard to enforce in a world in which streams of data are mixed and matched. And there is another tension between tighter data protection and more competition: not only have big companies greater means to comply with pricey privacy regulation, it also allows them to control data more tightly.



In time new technology, which goes beyond simple, easy-to-undo anonymisation, may ease such tensions. Bitmark, another startup, uses the same “blockchain” technology behind bitcoin, a digital currency, to keep track of who has accessed data. But legal innovation will be needed too, says Viktor Mayer-Schönberger of the University of Oxford. He and other data experts argue that not only the collection of data should be regulated but its use. Just as foodmakers are barred from using certain ingredients, online firms could be prohibited from using certain data or using them in such a way that could cause harm to an individual. This, he argues, would shift responsibility toward data collectors and data users who should be held accountable for how they manage data rather than relying on obtaining individual consent.

Such “use-based” regulation would be just as hard to police as the conventional rules of notice and consent which currently govern what data are collected and how they are used. It is also likely to worsen what some see as the third big challenge of the data economy in its current form: that some will benefit far more than others, both socially and geographically.

For personal data, at least, the current model seems barely sustainable. As data become more valuable and the data economy grows in importance, data refineries will make all the money. Those who generate the data may balk at an unequal exchange that only sees them getting free services. The first to point this out was Jaron Lanier, who also works for Microsoft Research, in his book “Who Owns the Future?”, published in 2014.

Mr Weyl, who collaborates with Mr Lanier and is writing a book about renewing liberal economics with Eric Posner of the University of Chicago, advances another version of this argument: ultimately, AI services are not provided by algorithms but by the people who generate the raw material. “Data is labour,” says Mr Weyl, who is working on a system to measure the value of individual data contributions to create a basis for a fairer exchange.

Data workers of the world, unite!

The problem, says Mr Weyl, is getting people to understand that their data have value and that they are due some compensation. “We need some sort of digital labour movement,” he says. It will take even more convincing to get the “siren servers”, as Mr Lanier calls the data giants, to change their ways, as they benefit handsomely from the status quo.

A more equal geographic distribution of the value extracted from data may be even more difficult to achieve. Currently, most big data refineries are based in America or are controlled by American firms. As the data economy progresses, this also hardly seems sustainable. Past skirmishes between America and Europe over privacy give a taste of things to come. In China draft regulations require firms to store all “critical data” they collect on servers based in the country. Conflicts over control of oil have scarred the world for decades. No one yet worries that wars will be fought over data. But the data economy has the same potential for confrontation.

This article appeared in the Briefing section of the print edition under the headline "Fuel of the future"