EuropeanVoice

Data: the new currency?

CONTENTS

Introduction

The hyperconnected world is producing data at an ever-increasing rate. In 2013, the global production of digital data reached four zettabytes, or four trillion gigabytes, more than double the amount generated in 2011.

According to one estimate, 90% of all data has been generated in the past two years. This data is coming from a variety of sources. People are using social media to post pictures and videos. More than 500 million photos are uploaded and shared every day while 200 hours of video are shared every minute. This is data that people are deliberately sharing about themselves. But this is only a fraction of the total amount of data that is generated and stored by all the digital technology with which people interact on a daily basis. This includes computers and internet-connected devices, but also data gatherers such as sensors, infrared cameras, closed-circuit TV cameras, radar and global-positioning satellites.

This huge increase in the generation, collection and storage of data offers enormous possibilities to amass useful information about people's behaviours and preferences. This information can be used to discern trends and patterns that can be used to design products and services to correspond more closely to consumers' preferences and also to improve efficiency in the provision of services such as transport and healthcare by matching resources more closely to demand.

Data analytics has been around for decades, if not centuries. But the massive amount of data being generated by the modern economy is a relatively new phenomenon that has been loosely denominated 'big data'. The term suffers from having a variety of definitions. For the sake of this report, we define big data as "data that is so large in volume, so diverse in variety or moving with such velocity that traditional modes of data capture and analysis are insufficient". 'Variety' refers to the collection of data from a range of different sources. This definition is often referred to as the 'three Vs'. McKinsey, a consultancy firm, defines big data as "datasets whose size is beyond the ability of typical database tools to capture, manage and analyse".

It is helpful to draw a distinction between big data and massive data. In general, massive data refers to large amounts of electronic information that may require supercomputers to handle the processing because of its size. In terms of format, massive data can be considered as the equivalent of spreadsheets or rows and columns of data – ie, in a uniform

What is happening in one day?

Email	Google queries	Pieces of content shared on Facebook
144.8bn	2.9bn	1bn

Tweets	App downloads on AppleStore	Pieces of content shared on Instagram
340m	67.5m	5.2m

Photos uploaded	Hours of video uploaded on YouTube	
500m	100,000	
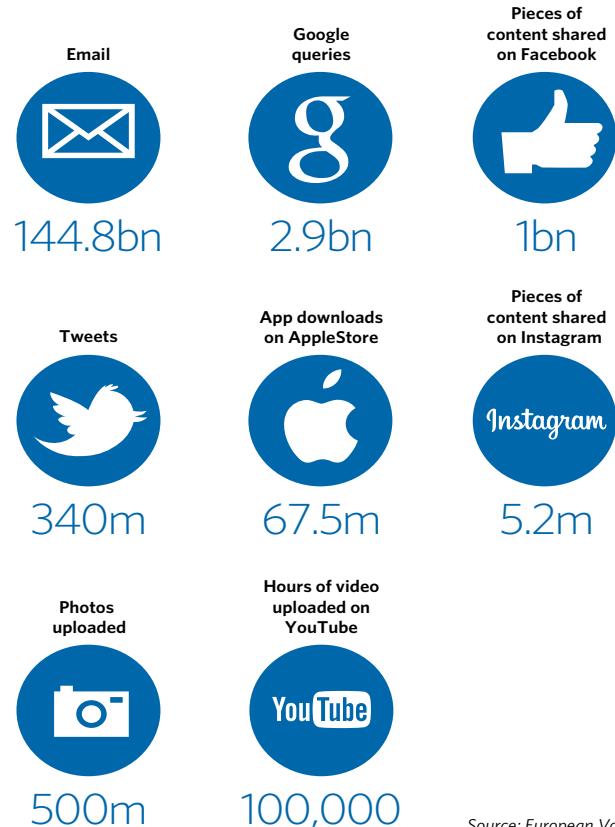
Source: European Voice

3

format. For big data, information is obtained from a range of sources in a variety of formats. This can include connected devices, sensors, cameras, etc. One of the characteristics of big data is that it relies on mixing different data sets in order to generate value. For example, Estonia has developed an effective traffic management system using mobile phone data. Whereas for massive data all datasets are processed, for big data part of the overall data is analysed for specific purposes. Velocity is a crucial defining element of big data because information, such as location data, needs to be processed in near to real time to provide useful information.

The collection of big data offers many business opportunities. But the beneficiaries are not exclusively businesses. Use of big data also has the potential to transform the planning and delivery of a range of public services including healthcare, disease control and transport planning. Amassing huge amounts of data about people's real behaviour in the public sphere, rather than what people say they do, can produce information that is used, for example, to tailor transport provision better to demand or to predict the spread of infectious diseases.

> Data that is so large in volume, so diverse in variety or moving with such velocity that traditional modes of data capture and analysis are insufficient

Definition of big data

Sandy Pentland, director of MIT's media lab, talks about a "data-driven world" in which the availability of fine-grained data about people's actual behaviour in social situations can be used to design better public policies. Pentland also sounds a note of caution: he warns that the ability to collect big data could be used to create an "incredibly invasive Big Brother". The threat to personal privacy needs to be addressed if the potential of big data is to be fully exploited. Otherwise people will refuse to share their data or will press governments to impose restrictions on data flow that could limit the uses of data. One of the major challenges of the big data era is how to ensure the protection of personal data in an environment where the value of the collected data derives precisely from it being combined with other datasets. Traditional forms of explicit consent to allow the use of personal data do not readily translate to an economy where data sets are used for purposes that were not envisaged at the time that consent was sought.

In this report we examine whether data can be the basis of a new transactional relationship between people and companies in which both sides benefit from new products and services and increased economic growth. The report includes recommendations to policymakers and industry on the actions needed to reap the benefits of big data while guarding against possible abuses.

Making the most of big data's business opportunities

Advertising and marketing companies have long been collecting and analysing data with the aim of gathering more information about consumer preferences and choices. What big data offers both the retail and services sector is the ability to aggregate data from different touch-points to obtain more granular knowledge. Data can come in structured forms, such as spreadsheets with columns and tables, but also more loosely, for instance via social media where individuals indicate their 'likes' and share them with their friends.

Using data from disparate datasets allows companies to target their marketing campaigns more precisely on consumers with particular likes, reducing the cost of campaigns and increasing their effectiveness.

Big data also has potential for generating cost savings in the healthcare sector. In the United States, rising healthcare costs have increased the importance of ensuring maximum efficiencies and increased the focus on payment-by-results. That requires the ability to gather data that links patient outcomes to forms of treatment. This information has to come from a variety of sources ranging from physicians and hospital records to healthcare insurers' databases.

Kaiser Permanente, a US healthcare management company, introduced a computer system to gather data about the details of cardiovascular disease sufferers and their treatment, including lifestyle advice and education. Data collected about patients and all forms of advice and treatment are collected into a single database that can be used by healthcare professionals to check the effectiveness of different types of intervention and make necessary adjustments. The programme lowered the risk of dying from cardiac disease and lowered cholesterol levels. The company saved up to €1 billion in reduced doctors' visits and laboratory tests.

Estimates by McKinsey say that use of big data could reduce healthcare spending in the US by €330bn-€450bn or 12%-17% of annual healthcare expenditure. Realising this potential will depend on a number of conditions, including ensuring that patient confidentiality rules are respected as datasets are combined. The transport sector generates large amounts of data including journey times, toll road data and congestion monitoring. It is already being used by application developers to help improve their travel plans. Logistics firms such as UPS are installing sensors in their vehicles to monitor journey times and improve fuel efficiency by avoiding congestion.

The next major development for the automotive sector will be the advent of Connected Vehicles, where data is shared automatically and in real time between the vehicle and road transport infrastructure. This will generate datasets that can be exploited through data analytics. Feeding information to intelligent transportation systems should lead to an overall improvement in traffic flow, fuel efficiency and pollution reduction. A test carried out in the Netherlands in 2012 generated estimates that up to 730 million tonnes of CO_2 emissions could be saved if applied to the country's entire fleet.

The value of big data is expected to continue growing as businesses increase their investment in big data technology. IDC, a market intelligence firm following the ICT industry, estimates that investment will continue to grow by 20% a year over the coming years. But realising its potential will require companies to acquire the necessary data-processing capacity to handle big data, including hiring qualified staff. For some organisations, there will be a challenge in switching to a data-driven model for supporting business decisions, rather than relying on past experience or a belief that an organisation instinctively knows what its market wants.

Growth of the big data market
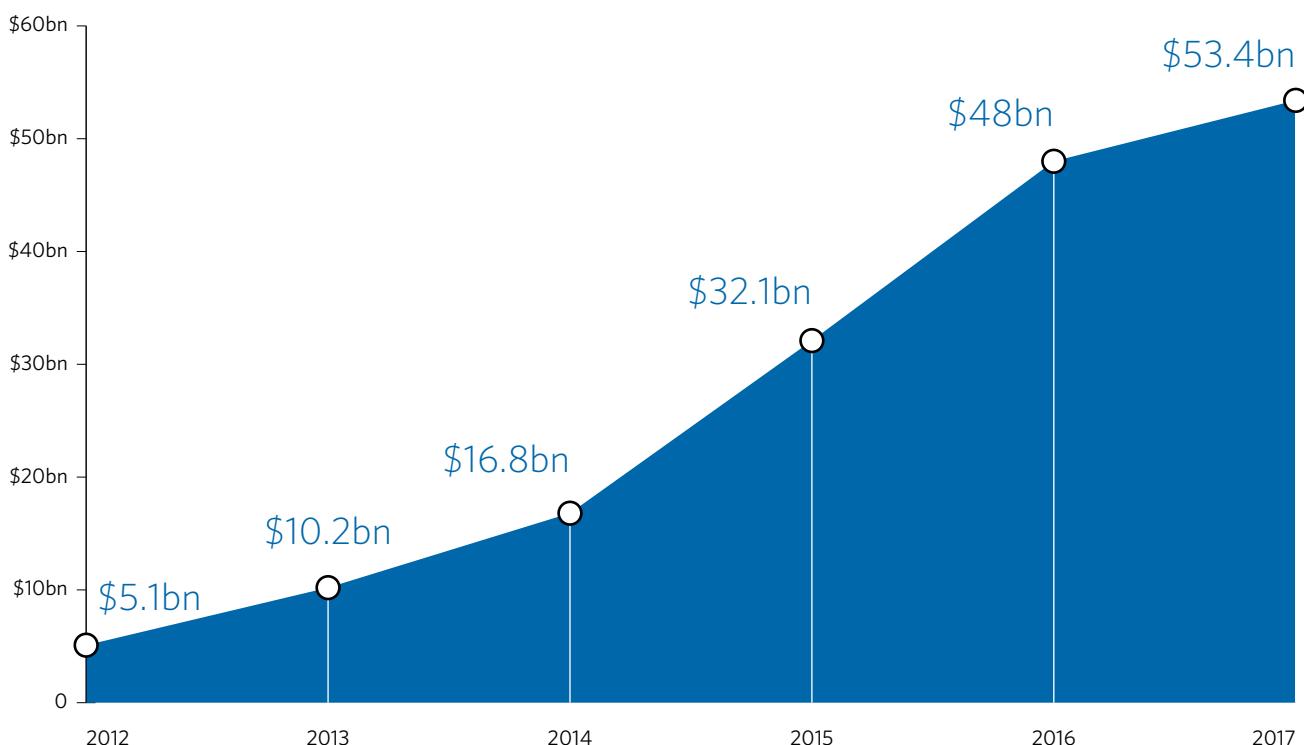
Source: Wkibon

Measuring data

One digit	A one-page text document	A song	A two-hour movie

1 octet	1,000 octets = 1 kilooctet	30 kilooctets	1,000 kilooctets = 1 megaoctet	5 megaoctets	1,000 megaoctets = 1 gigaoctet	1 gigaoctet
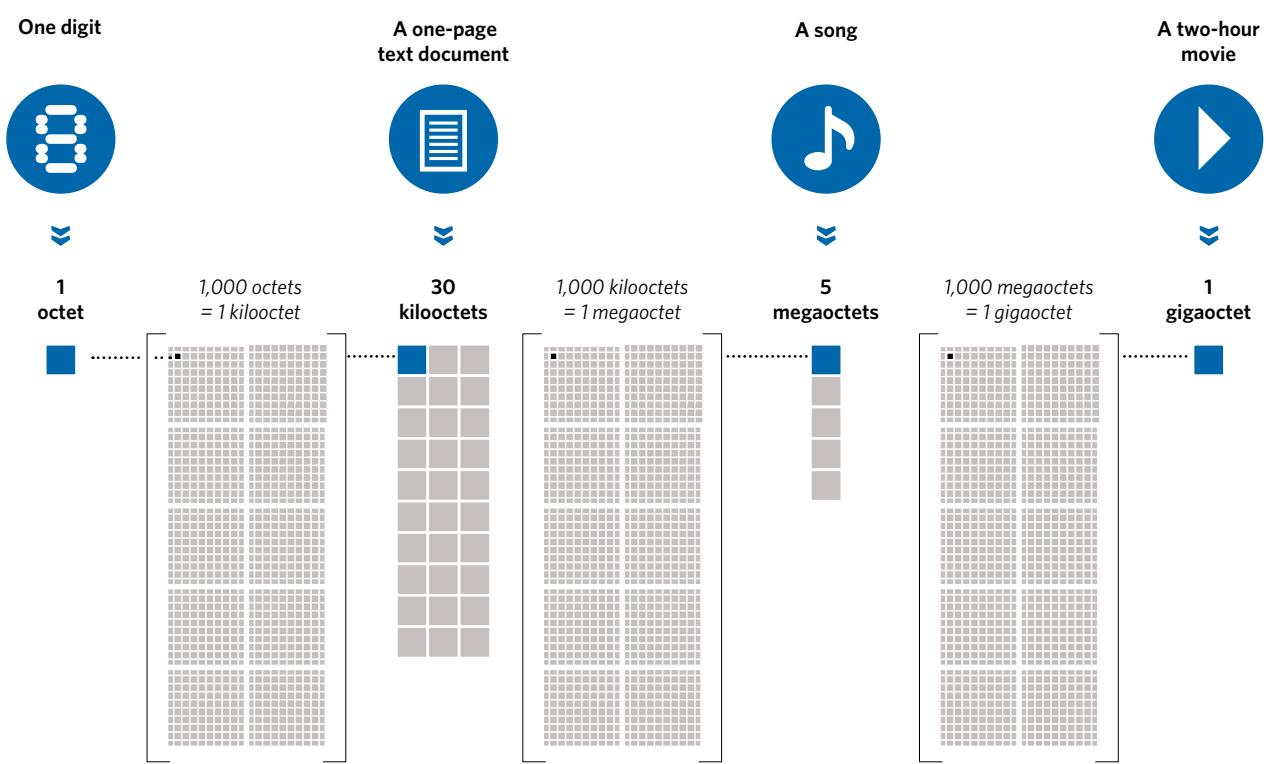
Extracting value from big data

Big data has the potential to improve the provision of essential public services such as healthcare. In 2008-09 the Harvard School of Public Health carried out a study in Kenya into how human population movements affected the spread of malaria. Scientists collected data from almost 15 million mobile phones and mapped calls and text messages to establish the length of trips and population movements. The data was anonymised and then compared with data about the spread of malaria from two malaria research projects. The study found that the spread of malaria was determined more by the movement of infected individuals than of infectious mosquitos.

The study identified areas where intervention would be most effective. The area with the highest rate of infection was Lake Victoria and it was identified as a source of infection for other regions. Thanks to the ability to target intervention, malaria death rates went down by 25% compared to 2000.

The San Francisco-based Global Viral Forecasting Initiative uses advanced data analysis based on a variety of internet resources to identify the locations and drivers of disease outbreaks before they become epidemics. This method has proved successful in predicting outbreaks up to a week before global bodies such as the World Health Organization.

The use of individuals' health data throws up one of the biggest challenges surrounding big data: the extent to which individuals are prepared to share their data and allow them to be passed to other organisations. Even where data is anonymised before being shared, it is relatively easy to link data back to specific individuals by comparing information from a range of databases. This has prompted fears that insurers could, for instance, use data to identify high-risk patients and raise their insurance premiums or introduce exclusion to their insurance cover. Procedures for anonymising data should be standardised to maximise individual privacy.

Opening up

In order to take full advantage of the potential of big data, it is essential that the vast amounts of data generated by organisations is made available in a usable form as open or 'liquid' data. Open or 'liquid' data needs to be machine readable, accessible to a broad audience at little or no cost and capable of being shared and distributed. According to a report produced by McKinsey in 2013, improving use of open data in seven sectors – education, transport, consumer products, electricity, oil and gas, healthcare and consumer finance – would produce $3 trillion (€2.2 trillion) in economic value.

The Organisation for Economic Co-operation and Development (OECD) estimated in a 2011 report that by fully exploiting public data, governments could cut administrative costs by 15%-20%, saving €150bn-€300bn. The data could come from public authorities, which are some of the most data-intensive organisations. Governments already have statistical departments to collate data about their population. The information derived from running mass-transit schemes and energy networks is a source of valuable public data.

There are numerous examples of the benefits of making such public data openly available, particularly in the transport sector. Transport data can be used by developers to create useful applications for passengers. Trafikverket, Sweden's public transport agency, provides third parties with real-time data on train departure and arrival times to allow passengers to make better travel plans. The European Commission has made increasing the availability of open data from public organisations one of its priorities. EU governments approved rules in April that would make all public sector information available for re-use, provided it is generally accessible and not personal. Under the rules, public sector bodies would be able to charge only the marginal cost of reproduction, provision and dissemination of the data. Data would be made available in machine-readable formats where possible.

The private sector also holds enormous amounts of data that could be used to develop new applications and services. Companies are traditionally reluctant to share data that they believe is commercially valuable or sensitive. There are a number of government initiatives to get companies to 'hand back' data by sharing what they have collected with consumers. The Midata initiative run by the Open Data Institute in London makes available data collected and held by companies. Firms from the energy, mobile phone and banking sectors made available data so that customers could learn more about their own behaviour and make more informed choices when purchasing goods online.

Making full use of the potential of open data will require efforts to agree common standards for the format of data and the rules for using it.

Open data in Europe

The Open Data Index assesses the state of open government data around the world. Score out of 1,000

	Score	Score as a percentage
UK	940	94.0%
Denmark	835	83.5%
The Netherlands	740	74.0%
Finland	700	70.0%
Sweden	670	67.0%
Bulgaria	520	52.0%
Malta	515	51.5%
Italy	515	51.5%
France	510	51.0%
Austria	505	50.5%
Portugal	495	49.5%
Slovenia	485	48.5%
Czech Republic	465	46.5%
Spain	460	46.0%
Ireland	460	46.0%
Croatia	445	44.5%
Poland	420	42.0%
Hungary	415	41.5%
Germany	410	41.0%
Greece	395	39.5%
Slovakia	375	37.5%
Romania	355	35.5%
Lithuania	320	32.0%
Belgium	265	26.5%
Cyprus	30	3.0%

Source: https://index.okfn.org

Trust and security

The era of big data poses major privacy challenges. When the age of computing started in the second half of the 20th century, information, such as tax or bank account details, was collected by a range of organisations, public and private, and held in information silos. Individuals gave their explicit consent to these organisations to collect and store this data but it was rarely shared with other parties. Then, with the explosion in computer technology, the internet and connected devices from the 1990s onwards, it became easier to collect and store data. Businesses emerged that made use of this data for sales analysis and marketing purposes. Western countries developed a legal framework to control the collection and use of personal data, encapsulated in the OECD's privacy guidelines from 1980.

The guidelines were first created at a time when the capture and use of data was less complex. Organisations collected data from individuals and stored it on computers. To protect data privacy, the guidelines focused on the use of collected data, limitations to its use and consent. This worked when data was held by organisations without being actively shared. Now, data can be generated without direct human interaction, such as through machine-to-machine transactions and GPS data. Companies have an incentive to obtain data from a range of sources to generate value. But traditional models for obtaining individuals' permission to use data, known as 'notice and consent', do not fit well with the new paradigm of huge data flows. It would take users an estimated 250 working hours to read all the terms and conditions that online companies ask individuals to approve in order to have access to their data. In reality, most users click 'yes' without knowing the terms for using their data.

The EU, which has some of the strictest data protection legislation in the world, is in the process of revising its rules, which date back to 1995. The rules require prior consent from users before data can be collected and strict rules on subsequent uses of that data. A fundamental element of the legislation is the right to be forgotten where users can ask to delete information held about them unless there is a public interest for that information to remain in the public domain. Companies warn that it can be difficult in practice to guarantee deletion as data can be held across a wide variety of databases. According to draft EU rules, companies that fail to respect data privacy could be fined of up to 5% of their annual turnover.

250

Estimated number of working hours that it would take to read all the terms and conditions that online companies ask individuals to approve in order to have access to their data

Confidence about internet transactions

Denmark and Sweden show high levels of confidence, with the lowest levels of confidence in Greece, Bulgaria and Portugal (figures from before Croatia joined the EU)   ▬ *Very confident*

	Denmark	Sweden	UK	Latvia	Malta	Ireland	Lithuania	Finland	Czech Rep	Estonia	Netherlands	Lux.	Cyprus	Poland	Slovenia	Austria	EU27
Red	3%	4%	10%	4%	15%	8%	11%	2%	6%	6%	4%	10%	24%	11%	13%	8%	12%
Orange	7%	4%	7%	14%	10%	9%	13%	10%	13%	12%	9%	11%	17%	14%	20%	16%	17%
Light blue	28%	35%	31%	31%	28%	37%	33%	49%	41%	42%	51%	44%	25%	40%	36%	49%	42%
Very confident	62%	56%	51%	50%	46%	43%	41%	39%	38%	38%	35%	32%	31%	31%	28%	27%	27%

8
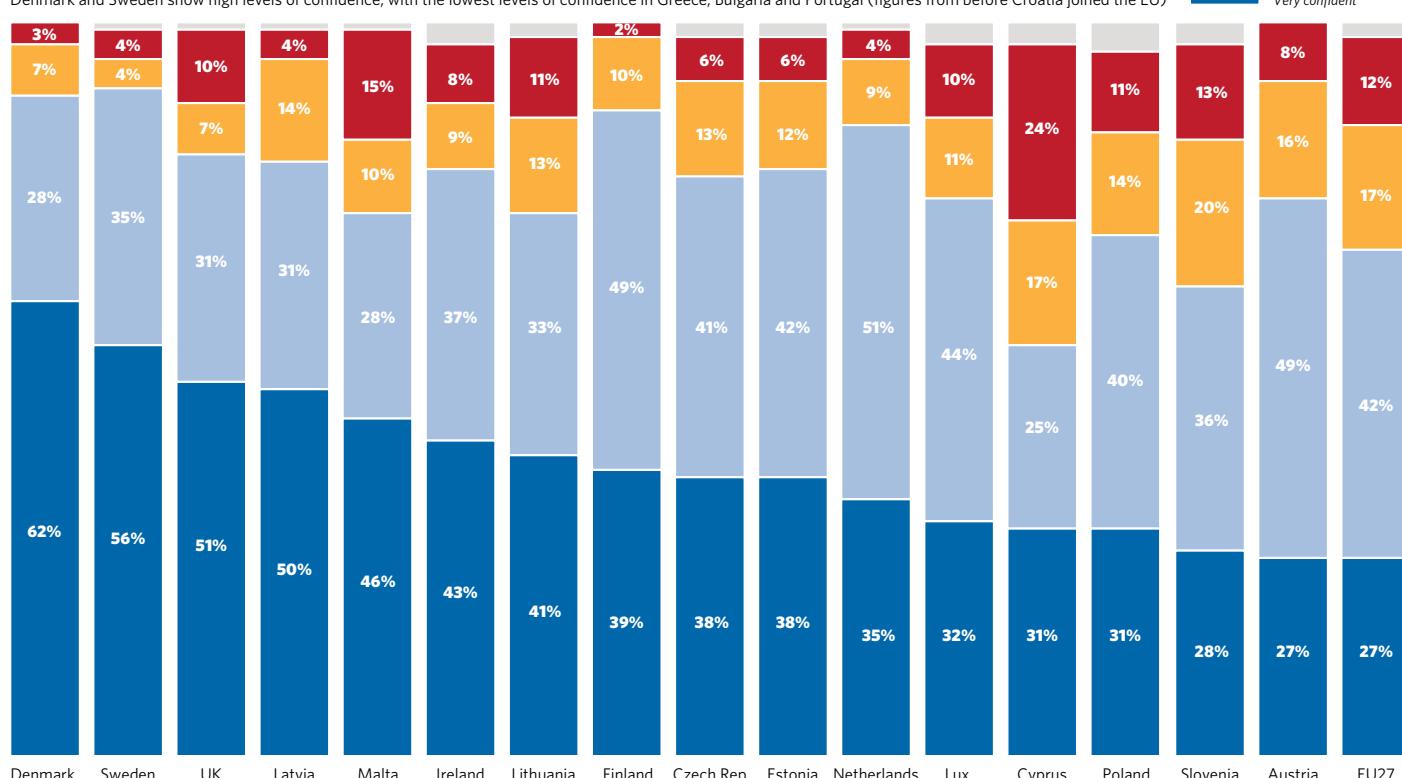
Building 'Fortress Europe'

As the EU's legislation imposes much stricter rules than those in other jurisdictions, such as the US, global companies warn that the EU is creating a data fortress blocking data flows to other parts of the world. The revelations by Edward Snowden, a former National Security Agency contractor, that the US government was spying on global telecommunications data have boosted support for even stronger data protection rules from MEPs.

> " We may be losing ground to the US when it comes to big data but we have a fundamental right to protect

German centre-right MEP Axel Voss

The European Parliament has also called for a suspension of safe harbour agreements with the US. These agreements allow EU bodies to share data with US companies provided they are certified as meeting EU data protection standards. The European Commission is examining ways to strengthen safeguards in these agreements to offer greater protection to EU citizens. The demands have prompted warnings that EU companies will be excluded from the business possibilities of processing global data. But MEPs are adamant that privacy protection must be maintained. "We may be losing ground to the US when it comes to big data but we have a fundamental right to protect. We should be careful with big data if we want to have privacy in the

future," says Axel Voss, a German centre-right MEP.

One approach to address the challenge of consent is to focus on anonymised data which, in theory, no longer identifies the individual who generated it. This is an approach that is attractive to companies. "The location of the customer is personal, but when you process it, you can anonymise it. Lots of data is user-generated but is not personal data after it is anonymised," argues Stefano Fratta of Telefónica. Opinions diverge about how robust data anonymisation can be. Some tests have shown that it is possible to identify a data subject from anonymised data using only relatively few pieces of information. A team of researchers at Harvard University was able to identify individuals from a genetics database by cross-referencing the information with other public databases. Using only three pieces of information, the team achieved an accuracy rate of 42% and this rose to 97% when first names or nicknames were added. Peter Hustinx, the European Data Protection Supervisor, says: "In reality it is now rare for data generated by user activity to be completely and irreversibly anonymised."

Clear definitions of rules for anonymisation processes for data processors will play a major role in boosting trust and ensuring that consumer rights are balanced with the potential to develop new services. In April, legal and technical experts on privacy and data protection from the EU and the US held a roundtable meeting to explore ways to bridge the gap between European and US legal systems of data privacy. The aim is to find globally-accepted privacy values to form the basis of interoperable solutions in both jurisdictions. The group, initiated by Jacob Kohnstamm, chairman of the Dutch data protection authority, will present a report in 2015. The answer may be to move away from trying to secure consent and towards a system of greater transparency in which companies are clearer about what they are using data for. Part of this is a privacy-by-design approach by which companies develop services with the highest level of data protection built in. One way of implementing this as a commercial service is by offering data protection seals, which would require organisations to meet clearly-defined standards for data protection. This idea is winning support from EU governments as a possible solution to the challenges of revising data protection rules. Privacy seals could indicate that an organisation was processing data in compliance with relevant aspects of the regulations, including core principles, privacy-by-design and security measures. The market for privacy seals is currently dominated by the US, and the EU would need to develop its own harmonised standards.

	Fairly confident	Not very confident	Not confident at all	Don't know

	Belgium	Slovakia	Germany	France	Hungary	Bulgaria	Romania	Italy	Portugal	Spain	Greece
Not confident at all	6%	14%	6%	15%	26%	27%	27%	14%	26%	18%	33%
Not very confident	15%	19%	18%	19%	15%	30%	18%	21%	30%	31%	29%
Fairly confident	54%	41%	51%	45%	31%	25%	36%	50%	32%	39%	29%
	24%	24%	23%	20%	16%	13%	12%	11%	10%	9%	8%
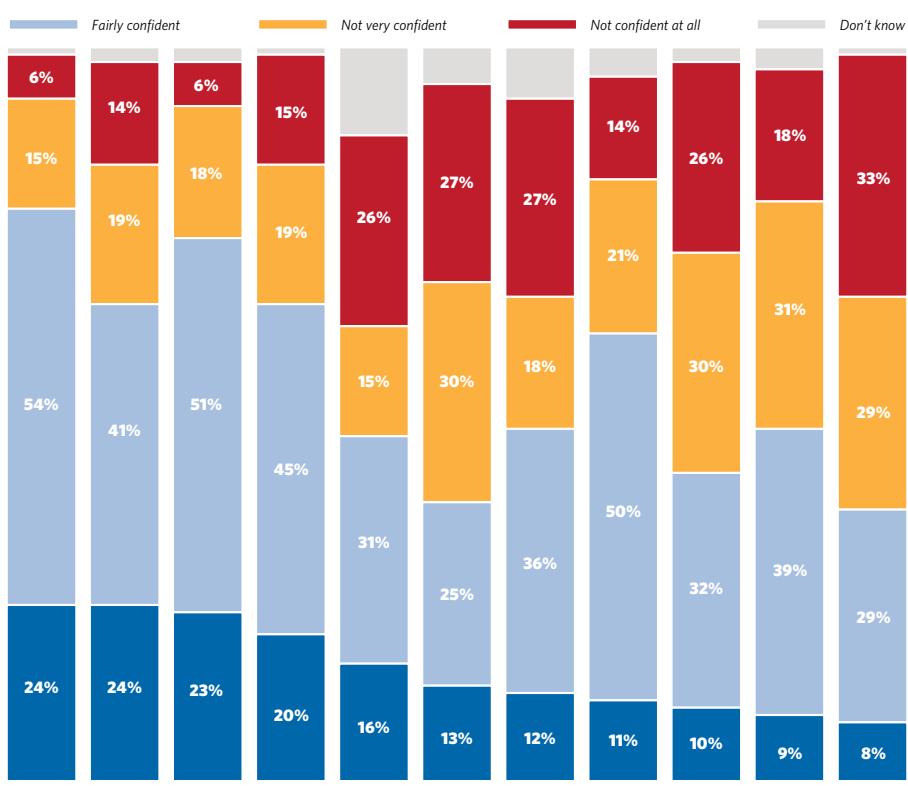
Source: European Commission

Cyber-security threats

Threats in the online sphere have the potential to inflict the greatest damage to trust and confidence. A Eurobarometer poll from November 2013 found that 37% of those surveyed were concerned about the misuse of their personal data in online activities; 35% were concerned about the security of online payments. Globally, victims' annual losses from cyber-crime activities are estimated at €290 billion a year, greater than the value of the illegal trade in marijuana, cocaine and heroin, while Europol puts the value of the global cyber-crime economy at $1 trillion (€739 billion).

Single incidents can have a huge cost. In 2011 Sony's Online Entertainment network was hacked, affecting 24.6 million users and breaching the security of 12,700 credit card holders. The incident is estimated to have cost the company €1bn-€2bn. In May this year, eBay, the online auction site, was hacked and the company warned its 145m customers to change their passwords to protect against fraud.

In addition to the economic damage from cyber-crime, the integrity of information networks can be attacked by criminals, enemy governments or malicious hackers. As so many essential services, such as energy, transport and finance depend on secure networks, attacks that impair the functioning of networks can have serious financial consequences. The World Economic Forum estimated in 2013 that there was a 10% likelihood of a major information security breakdown with a potential costs of €184 billion. In May this year, the website of the Belgian foreign ministry was hacked.

The EU launched a cyber-security strategy in early 2013 to address shortcomings in the current system. Not all member states had a dedicated cyber-security strategy in place; only a few member states were co-operating to tackle cross-border threats and many companies were failing to ensure adequate safeguards against cyber-attacks. The network information security directive, which has been agreed by the European Parliament but has yet to approved by member states, requires all member states to set up a national cyber-security strategy including Computer Emergency Response Teams (CERTS) to react to attacks and security breaches. National authorities will be expected to share information to improve the reaction to attacks, which are often targeted at several member states at the same time. Companies that operate network infrastructures would have to inform national authorities about attacks, whether or not a breach had taken place.

Effective cyber-security strategies are essential to address citizens' concerns about online security. Individuals will be wary of making credit card purchases or using online banking if they feel that they are vulnerable to data breaches or identity theft online.

Tackling cyber-crime depends on international co-operation

€184bn

World Economic Forum estimate of the potential costs of a major information security breakdown

among law enforcement bodies. The EU has launched the European Cyber Crime Centre within Europol to deal with international threats and co-operation with a particular focus on addressing organised crime and online fraud, child sexual exploitation and network security.

The EU has been co-operating closely with US law enforcement agencies to share information about online behaviour in order to identify terrorists or other criminals. It has negotiated a number of agreements to cover the terms for sharing information including the Terrorist Financing Tracking Programme (TFTP) to set out procedures for sharing data from SWIFT, an electronic bank transfer system. MEPs argue that the data-protection safeguards in this and other agreements are not strong enough and have argued that the agreement should be suspended and renegotiated.

The revelations by Edward Snowden about NSA surveillance of
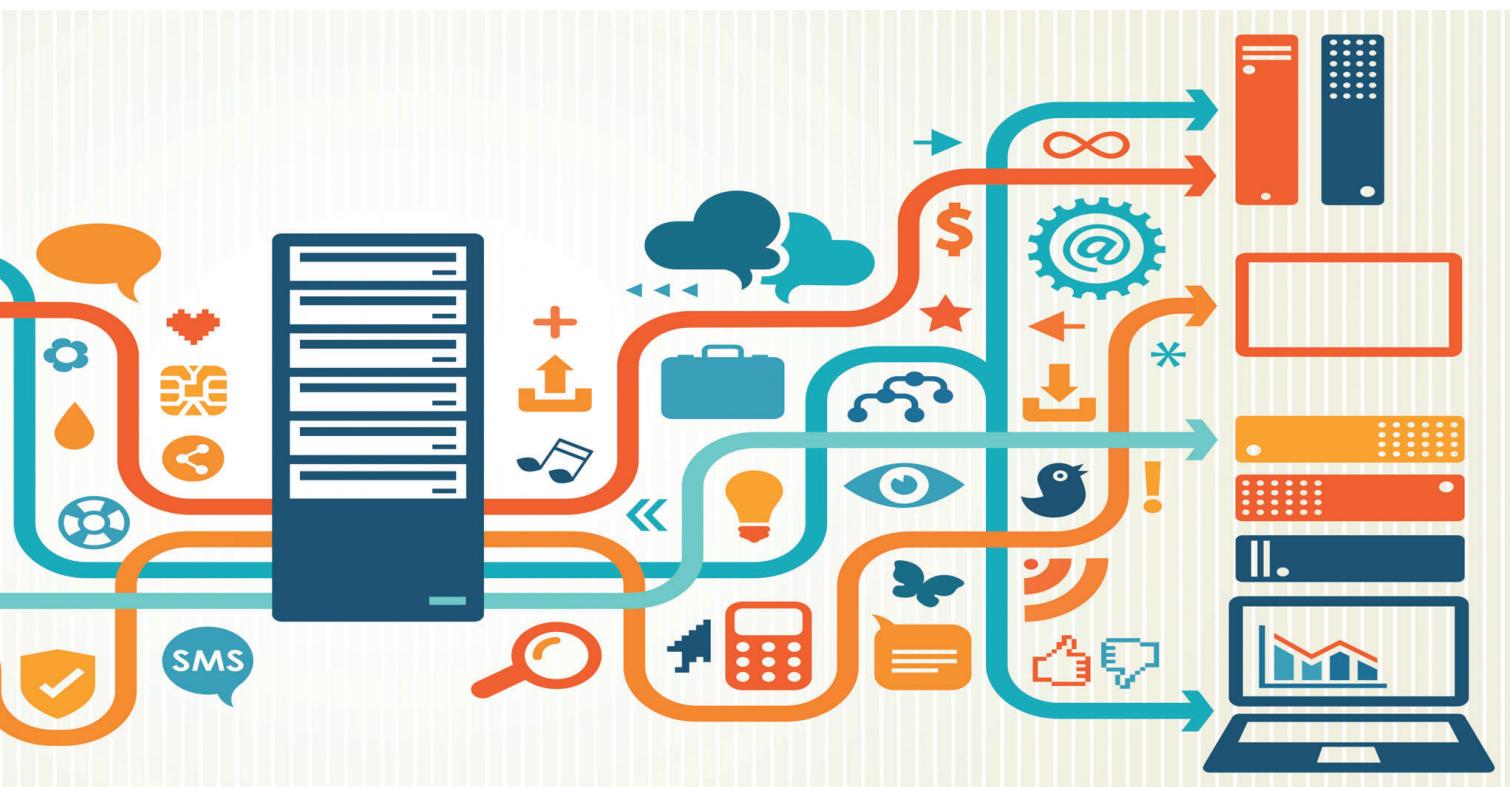
global telecommunications have increased calls for more robust data-protection rules. In February, the European Parliament's civil liberties committee voted to demand tougher data protection rules and said that the EU should block an agreement on a transatlantic trade and investment partnership (TTIP) with the US if the deal weakened EU data protection rules. The committee also called for the suspension of the 'safe harbour' agreement that facilitates data transfers from the EU to the US by setting out data protection standards which US companies have to meet in order to qualify to process data from the EU. The European Commission is currently renegotiating the terms of the 'safe harbour' agreement with the US.

Main concerns about cyber security

Losing personal data
40%

Security of online payments
38%

Prefer to conduct transaction in person
24%

Not receiving goods or services
19%

No concerns
21%
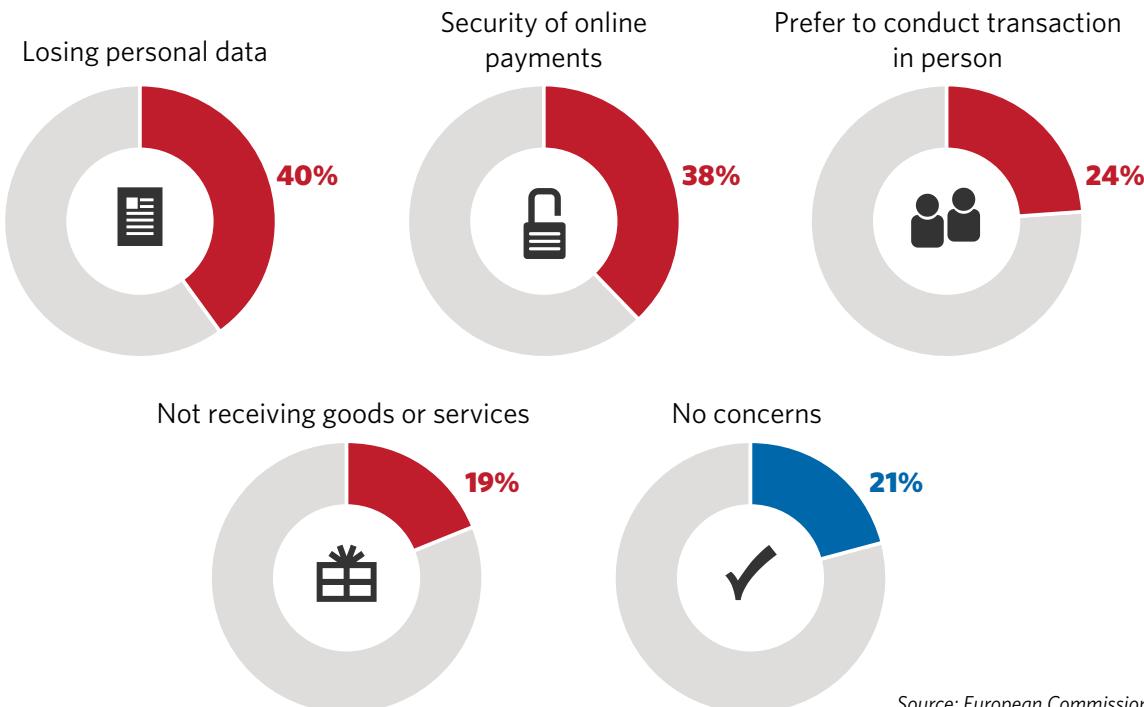
Source: European Commission

Conclusions and recommendations

The aim of this report is to review the issue of big data and to produce recommendations for policymakers and industry.

It is clear that the modern economy will increasingly depend on the use of the huge amounts of data that the connected world is producing at ever-increasing speed. The volume of data will increase further as more and more physical objects are connected to the internet, creating the internet of things.

> The economic potential of big data is growing, as long as companies have the right personnel and infrastructure to extract that value

In examining the specific nature of big data, its complexity and the challenging privacy issues it generates, it becomes clear that viewing data as a new currency only goes part of the way to capture the role of data in social interactions. Currencies function because the parties that use them agree about their underlying value: they are an intermediary for traded goods and services. Data, on the other hand, functions as a traded commodity itself. It has value to both parties in an exchange but there is an asymmetry in the value that different parties attribute to it and can extract from it. As this report has shown, the economic potential of big data is enormous and growing, as long as companies have the right personnel and infrastructure to extract that value. Consumers and individuals, on the other hand, may be unaware of the value of their data or find it difficult to retain some of that value for themselves. Some companies expect individuals to share their data for free in return for being able to use certain services. In the era of big data, where datasets are combined in increasingly complex ways, individuals quickly lose any share in the value that the use of their data creates.

If the benefits of big data are to develop in an equitable way, consumers will need to understand better how their data is collected and what it is used for in order to be empowered to retain some of the value of their data. This will require greater efforts by governments and educators to ensure that data literacy becomes a part of basic education. There are direct benefits to consumers from sharing data through improvements in healthcare services or shorter journey times. But these benefits have to be made more explicit so individuals know the

terms of the exchange. This exchange may take the form of a payment or the offer of new and better services that depend on consumers making their data available. There are examples where customers readily share sensitive data, such as banking and investment details, because they are confident that their data will be protected and used to provide better services. This is a model that could be extended to the big data era.

At the same time, businesses, individuals and policymakers have to be aware what big data is not: it is not an infallible guide to social trends. Like other forms of data collection and analysis, big

data shows correlation in datasets between behavioural patterns and trends. But to prove causal links between different phenomenon, intelligent analysis will still be needed.

To take an example, Google Flu Trends, an attempt to predict the spread of influenza based on online searches, produced results that diverged massively from actual disease tracking data. What the search had measured was queries about flu rather than cases. Taking big data findings as a reliable predictor of social behaviour would be to overstate the potential of big data and risk making errors in public policymaking.

Recommendations

1. **Improve digital literacy**
2. **Know the limits of big data**
3. **Open up data**
4. **Boost trust and security**
5. **Make redress real**
6. **Get the regulation right**
7. **Protect freedom of expression and human rights**
8. **Enable global data flows**

2

Know what big data can do
Big data is good at identifying correlations. It is not good at identifying causation. There is a need for intelligent analysis to assess what we can conclude from big data and what we cannot.

Action point: Clarify the limits of what big data can do.

3

Open data
Governments should take steps to ensure that data held by public authorities is made available as freely as possible and at minimum cost to users while ensuring that data privacy rules are respected. Companies should share the data they hold that can be used to develop new services for consumers.

Action point: Increase access to data held by public authorities and the private sector.

4

Trust and security
Consumers' and citizens' confidence in how individuals' information is used and protected in the era of big data is essential if the digital economy is to grow. Governments must ensure that fundamental principles of data privacy are respected while ensuring that rules are flexible enough to allow for innovation. Individuals will never understand the full complexity of the data ecosphere, so giving them control over their data through data vaults or privacy seals could play an important role in contributing to trust.

Action point: Establish rules that balance respect for data privacy with flexibility to support innovation.

1

Improve digital literacy
Current levels of data literacy are woefully inadequate. Both policymakers and individuals have a limited understanding of the digital economy. Governments, businesses and educators must work together to educate people about how the digital economy works and how data is used. At the heart of this initiative should be the aim of empowering citizens and equipping them with the knowledge to make better informed decisions.

Action point: Take action to increase data literacy.

13

5

Provide sanctions for abuse and ensure the right to redress

The financial potential of big data creates very strong incentives for some businesses to use personal data even when their use might conflict with data privacy. Measures will be needed to balance these powerful incentives with strengthened disincentives for abuse. This will require fines for companies that fail to protect personal data, and procedures for deleting or amending data that is inaccurate or in breach of data privacy laws.

Action point: Ensure ethical big data practices through sanctions for abuses and incentives to protect individuals' rights.

6

Get regulation right

Big data is qualitatively different from other forms of data as it depends on mixing data from different sources and applying insights from data analysis to a new set of individuals. A sophisticated regulatory approach, including competition issues, will be needed to deal with those actors in the data chain that are guilty of abuses. Clear rules about anonymisation will have a key contribution to make.

Action point: Develop an integrated approach to data protection, competition and consumer protection.

7

Protect freedom of expression and human rights

Governments should ensure that their national legal frameworks and actions are consistent with international human rights and standards of freedom of expression and privacy.

Action point: Ensure compliance with international human rights law and freedom of expression standards.

8

Ensure global data flows

Creating a 'Fortress Europe' for data will limit the potential for growth and innovation. The European Union and the United States must redouble efforts to work out common standards for the handling of personal data that allow for the free exchange of data. These standards should cover open data and anonymisation, among other issues.

Action point: Step up efforts to agree common EU-US standards for the handling of personal data.

Data generation

2010
1.2

2011
1.8

2012
2.8

2013
4

2020
40

Amount of data produced per year (in zettaocet)

90%

of data generated over the last two years is as much as in the previous history of the world
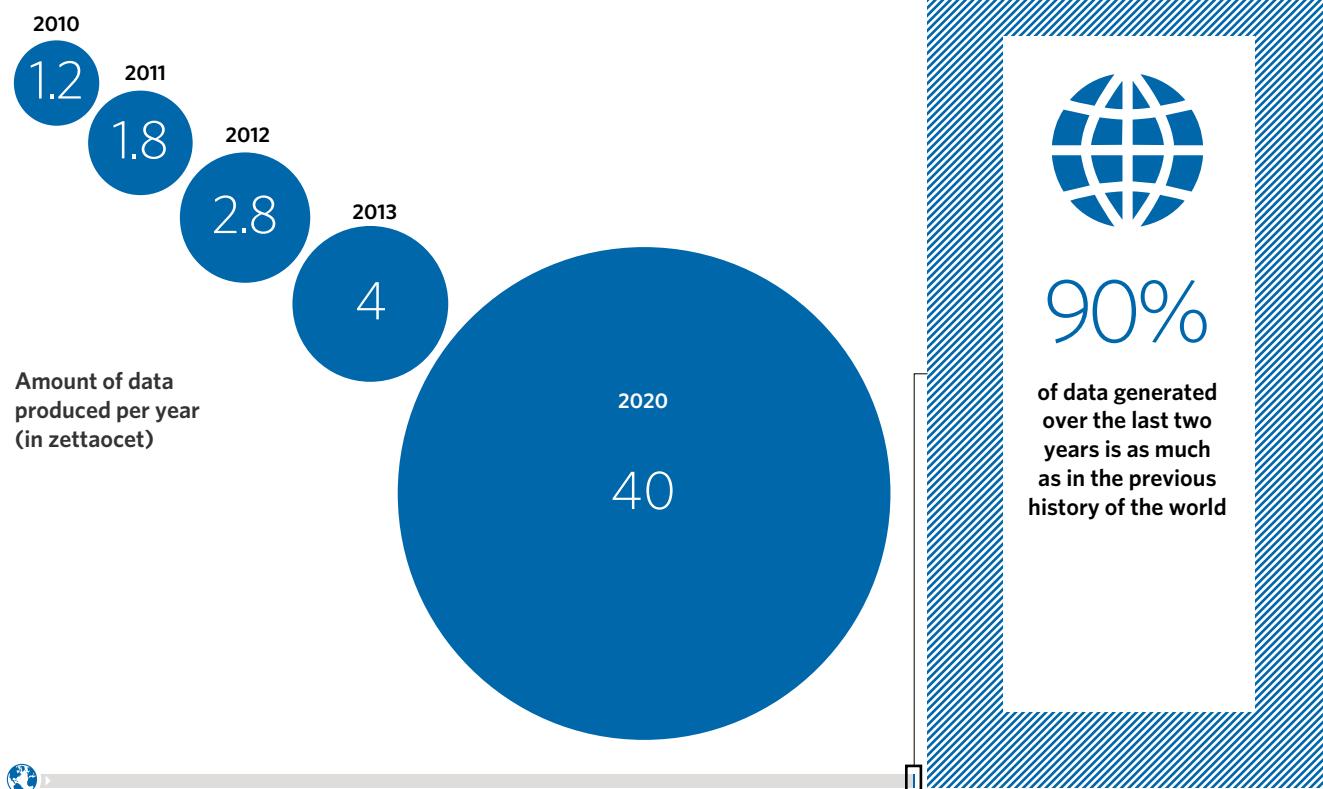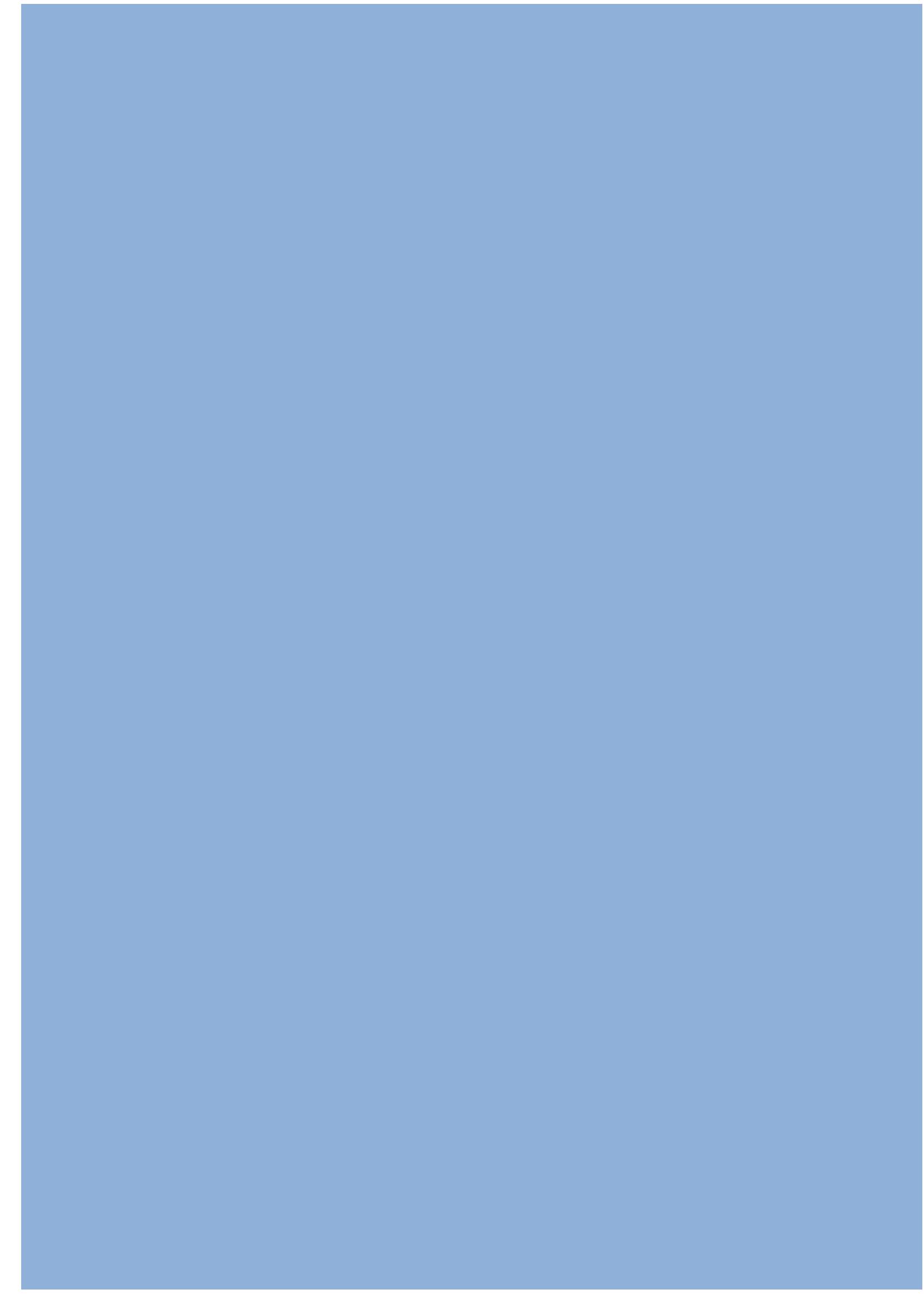
Source: European Voice

European**Voice**