



Enter the Data Economy

EU Policies for a Thriving Data Ecosystem

European Political
Strategy Centre

Data is rapidly becoming **the lifeblood of the global economy**. It represents a key new type of **economic asset**. Those that know how to use it have a decisive competitive advantage in this interconnected world, through raising performance, offering more user-centric products and services, fostering innovation – often leaving decades-old competitors behind.

As the world stands on the cusp of major new breakthrough technologies – Artificial Intelligence (AI), blockchain, robotics – advanced economies can reap significant benefits from embracing the data revolution. Recent research shows that even limited use of big data analytics solutions by the top 100 EU manufacturers could **boost EU economic growth by an additional 1.9% by 2020**.¹ And it is not only the manufacturing industry that stands to gain. Data analytics will soon be **indispensable to any economic activity** and **decision-making process, both public and private**.

Big data will be the backbone for successful and prosperous economies

By revolutionising business models, optimising production and decision-making, and enabling the development of innovative products and customised services, the data revolution can spur job creation and significantly enhance competitiveness and public service provision, to the ultimate benefit of EU businesses and citizens. Forgoing the opportunities of data would seriously harm the sustainability of the European economic and social model.

Better data governance is a prerequisite for a successful roll-out of big data strategies

Europe must ensure that the regulatory framework meets Europeans' needs and preferences, dispels uncertainty and increases trust for all players. This requires a good grasp of the complex and highly interconnected data ecosystem, of its main actors and of how it generates value.

With the GDPR as a backdrop, public intervention is also needed to govern non-personal data

To unleash the full potential of the data economy for businesses, top priority must be given to freeing up flows of data within the single market and with the rest of the world, promoting open data and transparency, increasing competition in the data value chain through enhanced interoperability and portability, and implementing clear liability rules.

Europe needs a far-sighted, proactive data strategy

Public intervention is required to incentivise the uptake of data services and reduce the cost of adopting data-driven decision-making. Targets and benchmarks should stimulate investment in digital technologies and big data analytics. This should be accompanied by investment in digital skills and organisational and financial support, particularly to small and medium enterprises. The latter may also entail a revision of state aid rules.

EPSC Strategic Notes are analytical papers on topics chosen by the President of the European Commission. They are produced by the European Political Strategy Centre (EPSC), the European Commission's in-house think tank.

Disclaimer

The views expressed in the EPSC Strategic Notes series are those of the authors and do not necessarily correspond to those of the European Commission.

The transition towards a data-driven economy **in Europe is trailing**, with **market players and public authorities reluctant or simply unable** to grasp new realities. To build a thriving data economy, Europe needs to dispel perceived uncertainties and overcome fragmented national environments. A sensible balance should be struck between data protection and consumer rights, on the one hand, and economic benefits and innovation on the other.

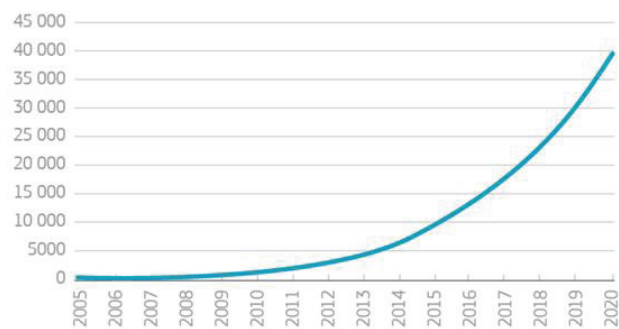
The EU's **General Data Protection Regulation (GDPR)**,² which regulates the processing and use of personal data in the EU, represents a first fundamental milestone to creating a **data-friendly environment where citizens and companies feel confident** that their privacy preferences are protected, while also safeguarding economic interests and innovation. Personal data represents one part of valuable business data; and businesses are now looking towards the EU to also ensure a level playing field and legal certainty with regard to the use of **non-personal data**³ so that **they can securely unleash the full potential of the digital economy and keep up with increasingly powerful competitors elsewhere**.

Big data as the bedrock of the future economy

The world is experiencing something akin to a data explosion. The amounts of information generated on a daily basis are not only vast; they are also growing at an exponential pace (Figure 1).

Figure 1: A global explosion of data

Worldwide data storage in exabytes



Source: International Data Corporation Digital University Study

Initially, this data fed through our fixed lines, mobile phones, laptops, tablets and other such devices. Increasingly, it is now also being transmitted by a whole range of other appliances, including connected cars, utility meters and consumer electronics – commonly referred to as the ‘Internet of Things’. The demand for digital technologies will continue to grow, as 61% of the world’s population has yet to go online.⁴ The Internet of Things is forecast to expand from a mere 4.6 billion devices in 2015 to more than 16 billion units in 2021.⁵

This move away from conventional, physical products towards complex, connected systems that combine sensors, software and digital user interfaces⁶ is creating a **value shift** that manufacturers simply cannot ignore. The real value is no longer in the product, as such, but in the opportunities it can offer to users in terms of accessing information and experiences.⁷ As businesses catch on to this trend, so too are consumers, with more and more buyers demanding products that are both personalised and connected across their various devices.

To a large extent, this value shift has been made possible as companies make use of the increasing amount of data they have at hand. The more data available, the better and more tailor-made the service they can provide. However, exploiting this high-volume, high-velocity and high-variety ‘big data’ to generate economic value not only requires technological developments, such as cloud computing and advanced analytics applications, but also **a new way of thinking, new skills sets and operating models**.⁸

The profound effect that data analytics is having on business is visible in the vast transformation underway in the car industry. Nine out of ten automotive executives believe in-vehicle and automation connectivity will disrupt their business model, while 80% believe that they will be challenged by new competitors in the field of connectivity and autonomous driving.⁹

Just as telling are figures on global trade, which show that data flows are emerging as a potential substitute for trade in physical goods, namely thanks to technologies such as 3D printing. Between 2008 and 2012, **world-wide cross-border trade in data increased by 49% while trade in goods or services rose by just by 2.4%**.¹⁰



Box 1: The sky is the limit: combining big data and space

The digital revolution, combined with decreasing costs of manufacturing and launching satellites, means space is becoming a powerful tool for collecting data at global and local scales.

The everyday life of Europeans has already become unthinkable without reliable satellites orbiting in space, as people across the continent make use of satellite positioning, navigation and timing services provided by Europe's Galileo satellites. The future potential is huge as it coincides with other developments for which navigation is a key enabler, such as connected vehicles, automated driving, drones, etc.

These developments not only offer opportunities to private companies, they also provide for tools to address emerging societal challenges and contribute to better policymaking. For example, Europe's Earth observation system Copernicus can help us to understand changes to our planet, such as rising sea levels, ice concentration or changing atmospheric conditions. Thanks to these insights, it is possible to monitor the state of extractive industries or vital infrastructures like power grids, improve maritime surveillance, prevent illegal fishing, better manage natural disasters, obtain early warning of possible migrant flows, etc.

It is now a matter of capitalising on these public investments by better linking space activities with all sectors of the economy and transforming the raw data into products and services that have an economic added value through big data analytics.

Data as the new driver of productivity, jobs and innovation

The economic benefits of using big data are underscored in numerous studies. Firms that adopt data-driven decision-making have been found to have a **5-6% higher output and productivity**.¹¹

Companies can use big data analytics to help them develop new products and services, to re-engineer their business processes and better manage their supply chains, to strengthen fraud detection, to improve security and risk management and to gain clearer insights into customer needs.

Latest studies at EU level estimate that 100 000 new data-related jobs will be created in Europe by 2020,¹² while the introduction of big data in the top 100 EU manufacturers *alone* could lead to savings worth 425 billion euro, representing a GDP increase of 206 billion euro or 1.9% over the same period.¹³ Another study finds that big data analytics solutions have the potential to unlock an additional 270 billion euro in economic benefits for the UK over the period 2015-2020.¹⁴ This is equivalent to an average of 2.0% of UK GDP *per year* (Figure 2).

There are thus huge opportunities to be reaped from 'upgrading' Europe's traditional industries to take full advantage of the digital transition.

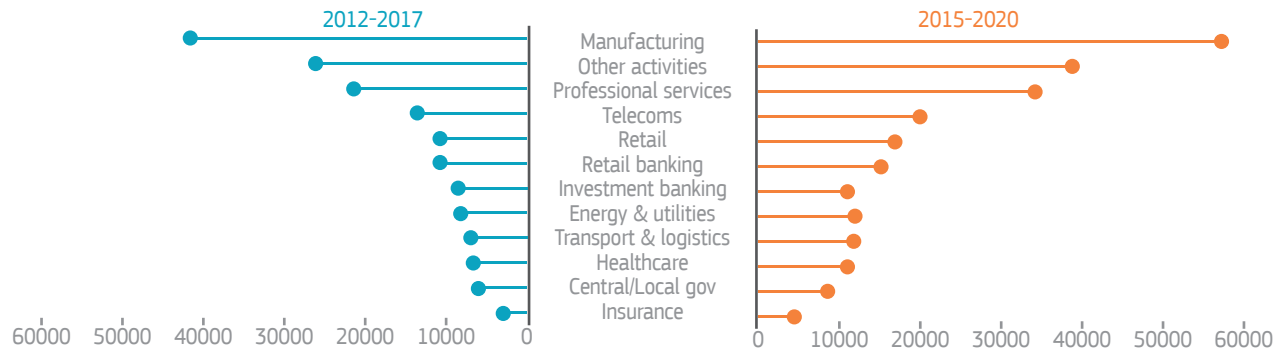
Many companies have already understood the potential benefits of implementing data-driven decision-making, and are investing rapidly in big data technologies and

services. The **global market for big data**-related hardware, software and professional services (such as data-centre computing, networking, storage, information management or analytics) is booming and is forecast to reach **43.7 billion euro by 2019** – ten times more than in 2010 (Figure 3).¹⁵

Yet, concerns about the disruptive economic effects or the potential misuse of big data analytics are real and need to be taken seriously by public authorities. Business and societal transformations rarely come without side-effects. One example is the emergence of Artificial Intelligence (AI), which is inherently based on automated learning processes enabled by big data analytics. On the one hand, machine learning has the potential to disrupt current working patterns, leading to job displacement or even destruction as intelligent devices can replace human labour. This will require close monitoring, as well as public policy intervention to support those laid-off, with the goal of reintegration in the labour market. On the other hand, important ethical questions arise with the use of AI for predictive analysis as this could lead to an increased risk of discrimination. For example, data analytics could be used to predict the likelihood of pregnancy among job-seekers or to assess the future health prospects of a job applicant. Also, services aimed at making everyday life 'easier' could end up reshaping human interactions – for example, through 'predictive chat apps' that suggest predetermined emotional responses on the basis of the content shared by users. The novel risks emerging with the onset of the data revolution should not be ignored as their social impact can be profound.

Figure 2: Big data, big benefits

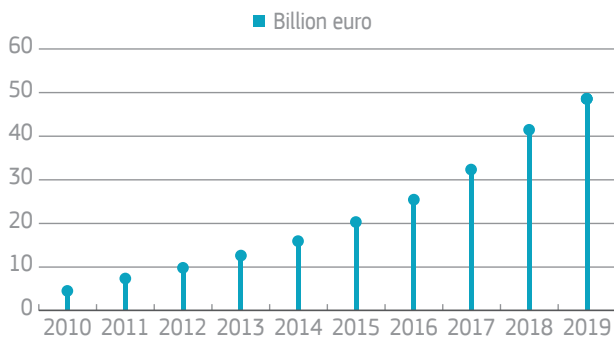
Cumulative economic benefits of big data analytics to UK industry, million pounds



Source: SAS & Centre for Economics and Business Research Ltd, 2016

Figure 3: Impressive growth of big data market

Worldwide big data technology and services forecast, billion euro



Source: International Data Corporation Digital University Study

While the GDPR establishes an important and one-of-its-kind regulatory framework to meet those concerns with regard to personal data, the fact remains that many machine-generated data are not personal and their processing therefore does not fall within the scope of the GDPR. From an economic point of view, and given the surge in available data, this limbo should raise concerns and calls for action.

Europe lagging behind in embracing the digital and data revolution

Big data and the Internet of Things are at the heart of a new industrial revolution, changing the landscape of our economies and boosting global productivity on a similar scale as the emergence of steam power during the first industrial revolution.

And yet, **in Europe, too few companies are embracing digitisation:** In 2015, only one in five European companies displayed a high or very high digital intensity,¹⁶ while only 6% of ICT and professional services companies were making strategic and intense

use of data.¹⁷ Data specialists accounted for (far) less than 1% of total employment in most Member States.¹⁸ And while the lack of supply of data-savvy workforce is a problem that also affects other technologically advanced countries such as the United States, Europe also appears to be lagging behind on big data infrastructure. More than 50% of all data centres in OECD countries are located in the United States (Figure 4).

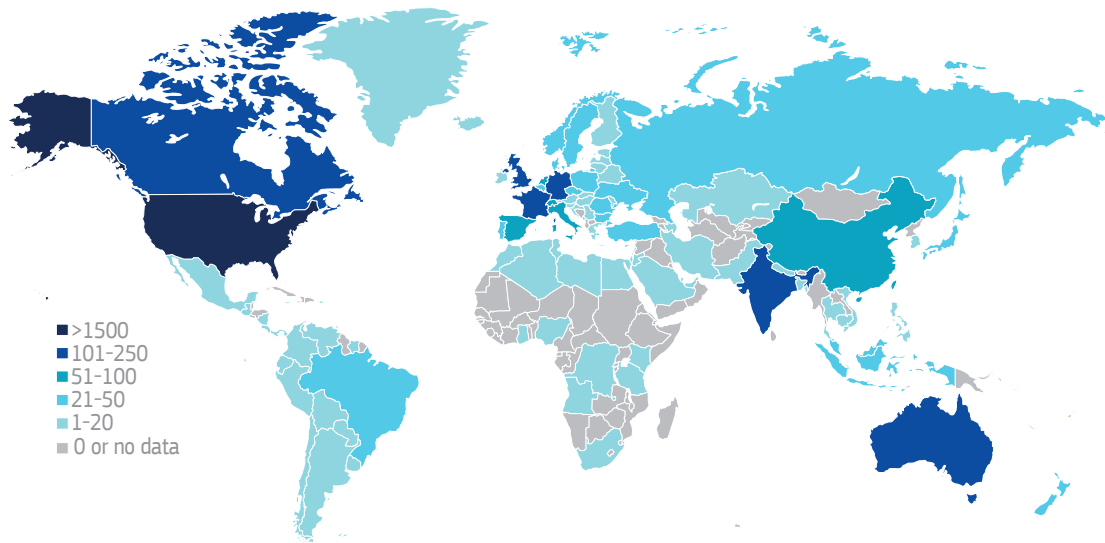
These are **worrying signals:** in the face of an ever increasing global competition, the European economy cannot afford to miss the data revolution opportunity and lose its competitive edge. With limited access to data and data analytics, European companies will not be able to compete in global markets and small and medium enterprises (SMEs) and emerging companies are the ones set to lose the most. A McKinsey study predicts that China is likely become the hotbed of the ‘car data revolution’ as Chinese customers tend to be much more open to information sharing than Europeans or Americans.¹⁹

Equipping European businesses with the means and tools to capture, process, store and analyse big data and generate value from it is a means of securing future wealth and prosperity.

Moreover, Europe’s public administrations have to embrace the new paradigm, exploiting digital opportunities to gain efficiency, reduce costs and expand the range of services provided to citizens and businesses within the new framework set by European law. They also have to learn how to build local data ecosystems to foster entrepreneurship and social innovation, in close cooperation with civil society and citizens’ associations. By transforming the way in which statistics are produced and used – and by increasing timeliness and granularity of key data – the data revolution can dramatically enhance the quality of policymaking and increase the accountability of public authorities, helping to overcome the growing distance between public institutions and citizens.

Figure 4: EU slow to invest in big data infrastructure

Number of data centres

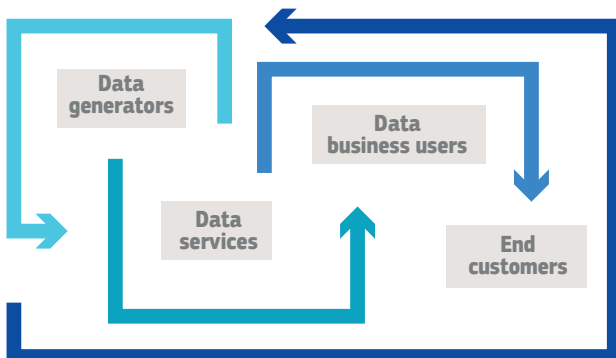


Source: <http://www.datacentermap.com/>

Step 1: Understanding value creation in the data ecosystem

Broadly speaking, there are four categories of actors in the data ecosystem (Figure 5).

Figure 5: Main actors of the data ecosystem



The data generators

These are the **primary source** of data: users browsing the internet; consumers paying with their credit card; location data from GPS readings; a smart-fridge's function to reduce energy consumption; or simply rain

on a summer's day in Amsterdam. Any action or non-action by humans or non-humans conveys information that can potentially be collected and aggregated for meaningful data analysis.

Data can be generated *passively*, when the subjects generating data are not necessarily aware that their action is conveying information being captured by another entity (e.g. when a driver uses a mapping application). Or they can be generated *actively*, for example if a user fills in her email address to subscribe to the services offered by a social network. This distinction is relevant, among other aspects, to determine whether the data generator has given its consent to the processing of the data and what is the purpose of the data collection (and thus whether it is compatible for further processing – see Box 3).

Since the information created is valuable to data services, **data generators may be able to benefit by 'monetising' this information** (e.g. in exchange for a price discount on a related product or service). Most of the time, transactions remain implicit. One classic example is that of the advertisement business model, which allows consumers to use services for free (e.g. a search engine or an online news service), while using the information provided for targeted advertising purposes.

Conversely, data generators may be **exposed to risks, such as privacy and security breaches** – even when

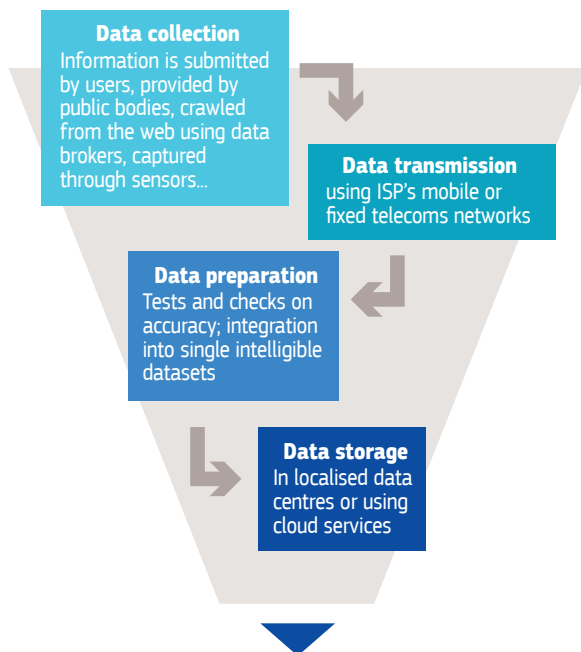
data is safely stored. Such risks can be mitigated by *encryption* or *pseudonymisation* of datasets. Nonetheless, there are increasing concerns around the possibility to ‘de-anonymise’ data, by cross-referencing it with other sources of data to re-identify the anonymous data source. A recent UK Parliamentary report has recommended making de-anonymising data a criminal offence.²⁰

The GDPR takes account of these risks, to the extent that the generated data is considered personal or that the natural person at the origin of the data can be identified despite pseudonymisation. Controllers and processors of personal data will have the obligation to implement ‘appropriate technical and organisational measures to ensure a level of security appropriate to the risk’²¹ – such as pseudonymisation and encryption. Failing this, they risk incurring impressive fines.

Data services

The information underpinning what we call ‘data’ is just a by-product of reality. The actual creation of value only occurs when that information is processed and analysed (Figure 6). This includes collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.²²

Figure 6: The data value creation chain



Big data analytics: statistical analysis, visualisation and other reporting and analytical techniques, including automated analytical model building known as ‘machine learning’

Bottlenecks may arise **at any level of this data value creation chain**, reducing the generated value to the detriment of the different data service providers themselves, as well as the data-users and downstream consumers (particularly when data is not easily transferred from one service to another). This can happen when competition among service providers is limited: If they enjoy sufficiently high market power, internet service providers can, for instance, limit data traffic, while infrastructure service providers can practice vendor lock-in by using proprietary technologies that make it difficult or impossible to transition to a competitor’s product or service, implying a decrease in customer choice.

In fact, there is a strong tendency of **market concentration** in some segments of the data industry, particularly in areas that feature very strong network externalities and economies of scale. Simply put, the more data is fed into the analysis, the better and more efficient the service becomes. As Shapiro and Varian put it: ‘Positive feedback makes the strong get stronger and the weak get weaker, leading to extreme outcomes’²³ – for example: the more users use a web search engine, the more feedback is received and the stronger is the ability of the search engine to return results responding to what users are really looking for – ultimately reinforcing the search engine’s market position.

Bottlenecks can also be **created by regulatory interventions** that limit accessibility to data and its movement, and create high compliance costs for service providers. For example, data localisation conditions requiring service providers to process data in specific geographical areas can prevent businesses from delocalising to areas where data processing and administrative costs are lower.

Data business users

These are the companies and public administrations using the *outcome* of data analytics to improve performance, ranging from incremental to more disruptive changes. This can encompass enhancing internal monitoring, tailoring marketing strategies, cutting processing costs, deploying entirely new products, or even fundamentally overhauling business models.²⁴ It can be expected that, as companies gain experience with data analytics and see the potential gains, the balance between **incremental and disruptive changes** will tilt towards the latter.²⁵

Nonetheless, in order to embed big data analytics within their organisation, companies and public administrations need to overcome a number of **barriers**, the most relevant being **lack of awareness** of the functioning

and the potential benefits of data analytics, resistance to and fear of potential organisational change entailed by big data analytics, **absence of available technological means**, particularly as regards the ability to integrate and manage large datasets, shortage of skilled, data-savvy **staff** and lack of **financial means** to make the technological and staffing investments. This is a particular challenge for SMEs.

The take-up of big data analytics can also be hampered by lack of clarity in the regulatory framework regarding privacy, data protection and security.²⁶ Here, progress has been made in Europe with the adoption of the GDPR as regards the processing of personal data. However, in order to stimulate the growth of the data economy, a common EU framework facilitating access to non-personal data, such as some machine-generated data, is needed to overcome national fragmentation and provide comfort to innovators. Moreover, the role of regulators and policymakers must also be to proactively support and incentivise a data-friendly corporate culture, for instance by setting benchmarks to be achieved (i.e. a digital investment target comparable to the 3% Research and Development (R&D) target) or considering tax credits (as has been done in several Member States to drive R&D investments).

End customers

The last group of affected actors are the downstream buyers: the consumers, business customers or citizens dealing with companies or public administrations that have implemented big data analytics within their organisation. The most straightforward effect of data analytics for this category consists of potential price reductions, better quality services or products, including the possibility of accessing novel products that would not exist without data analytics. In addition, one might benefit from better tailored offers, powered by more accurate marketing strategies. Applications that map traffic in real time and allow drivers to select the best route to avoid road congestion are just one example of this. Such applications rely on a large number of drivers sharing information about traffic while driving. Without big data no such service could exist.

Yet these advantages can also come with downsides: Because the end customers are often also data generators themselves, they are exposed to security and privacy risks. Moreover, **they are more vulnerable to exploitative pricing practices**, where sellers use data analytics to extract a more accurate profile of the buyer and accordingly charge a price that is closer to the buyer's willingness to pay. While 'price discrimination' as such does not necessarily imply inefficient outcomes from a welfare perspective²⁷, it typically implies a shift of surplus from customers to seller. **Such price discrimination is not necessarily linked to the processing of personal data and might thus occur in the limbo outside of the GDPR protection.** For example, medical insurance companies could use statistical inferences based on ever-larger non-personalised datasets to increase premiums in regions where certain diseases have higher incidence.

Table 1: Summary of benefits and effects for actors in the data value chain

ACTORS	BENEFITS	RISKS / BOTTLENECKS
Data generators	<ul style="list-style-type: none"> Ability to use own data as a commodity 	<ul style="list-style-type: none"> Security breaches Privacy breaches
Data services	<ul style="list-style-type: none"> Profit generation 	<ul style="list-style-type: none"> Bottlenecks in data flows Compliance costs of regulatory framework
Data business users	<ul style="list-style-type: none"> Increased productivity Incremental and disruptive innovation 	<ul style="list-style-type: none"> Investment costs vs regulatory uncertainty Resistance or incapacity to implement big data analytics
End customers	<ul style="list-style-type: none"> Lower prices and higher quality Tailored offers New services/goods 	<ul style="list-style-type: none"> Price discrimination Security and privacy breaches

It is important to note that the classification in Table 1 is made only for analytical purposes. In reality, it would be hard to observe actors that only fit one of the categories. Most commonly, categories overlap as data generators are also end customers or companies that use data analytics generate data themselves and may run their own data analytics services.

Step 2: Building a data-friendly regulatory framework

The European Commission has recognised the importance of advancing the data economy and prioritised several actions in its *Digital Single Market (DSM) strategy*²⁸. One of the key building blocks of this strategy is the GDPR. Market forces alone cannot be relied upon – essentially because of **two main failures** in data markets.

First, complexity of data dynamics, lack of transparency and lack of competition at control points imply a risk of harm to businesses and citizens through **privacy offences, security breaches or other harmful practices, such as de-anonymisation or price discrimination**. Data generators and customers are often unable to appreciate and anticipate the extent to which they are exposed to potential harm because of the information they provide (possibly unawares). And, even if they were, they might still play along because few alternatives exist.

Second, market forces alone are unlikely to lead to a socially optimal use of data because they are unlikely to make full use of the potential spill-over effects of data. Indeed, the ‘non-rival’ nature of data means it can be used multiple times and/or by multiple users without depleting its value.²⁹ Economic theory suggests that, in principle, any data collected has the potential to be usefully analysed for entirely new purposes and to bring promising new insights that could generate substantial benefits. For instance, data that has been collected for marketing purposes could, in principle, be re-used to improve research in healthcare at no or minimal additional cost for the data collector (in the EU, in the case of personal data, this will require user consent or other lawful grounds explicitly mentioned in the GDPR). If these potential spill-over effects are not taken into account, data analytics produced in the economy will, at best, match companies’ private benefits and not the social value. This leads to a risk of undersupply of data analysis i.e. less data is collected, processed and analysed compared to what would be socially optimal.³⁰

Public intervention must address these market failures, but it must be carefully balanced and based on consistent principles targeted at maximising social welfare. Regulators face **difficult choices** in order to create a vibrant digital economy and society while guaranteeing high data protection standards.

At the same time, there is a clear **first-mover advantage** for Europe to take the lead on setting regulatory standards governing the data economy as a whole, as this will ensure that they are aligned with European needs and preferences. A first clear and important step in this direction has been taken with the GDPR (Box 3).

Box 2: Harnessing big data to improve healthcare performance

An area where major breakthroughs are made possible through big data analytics is the health sector, although this is clearly also an area where personal data is particularly sensitive.³¹

Big data can be used to predict the likelihood of developing a rare disease, to monitor adverse events related to pharmaceuticals and medical devices, to provide faster health diagnostics (e.g. using information provided from wearable devices that track your exercises and progress), to offer more personalised healthcare solutions or to dramatically enhance the effectiveness of treatments.

For example, the **EU Advanced Drug Reporting initiative (EU-ADR)**³² currently automates the analysis of data stored in anonymised **electronic healthcare records** of over 30 million citizens in selected Member States to monitor the effects and safety of infrequently used drugs. The mobilisation of big data makes it possible to detect nuances in sub-populations that are rare and would otherwise not be readily apparent in smaller samples.³³

Machine-learning algorithms have also been used by researchers from Carnegie Mellon University in Pittsburgh to work on cardiac or respiratory arrest to predict, up to four hours before the event, whether a patient will go into arrest. Their predictions proved correct in two-thirds of the cases, against a 30% accuracy without big data analytics.³⁴

With Europe’s particular demographic profile, as a region with a fast-ageing population, personalised healthcare holds the potential for more targeted medical interventions, with the prospect of helping people live healthy lives for longer, while enabling cost savings for social expenditure, without compromising quality of care.

Delays in action on building a friendly legal framework covering non-personal data would entail the risk that global regulatory standards emerge elsewhere, forcing Europeans to adapt, e.g. on issues of access, interoperability or liability. In July 2016, the Chinese Government announced that it would soon take important legislative measures in the field of cyber-security, data handling and online activities. In particular,

a proposal on coding property rights for data in the context of a redrafting of the Chinese Civil Code is currently under discussion.³⁵

The bottom line is that if Europe does not articulate a vision – and accompanying standards and rules – for non-personal (big) data, others will do it for us.

Box 3: General Data Protection Regulation: A game changer for Europe's data economy?

Article 8 of the Charter of Fundamental Rights provides the right of protection of personal data, specifying that it 'must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified'.

The General Data Protection Regulation (GDPR),³⁶ which will enter into force on 25 May 2018, lays down the rules to protect this right. From this date onwards, national legislation on data protection and data flows (concerning personal data) will become obsolete and replaced by the harmonised and unified regime of the GDPR. Many see this as a major milestone to overcome Europe's heavily fragmented data privacy regime, which some argue has been a major impediment to building up sizeable companies in this space.

The GDPR defines personal data as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (Article 4(1) GDPR).³⁷

At the same time, the GDPR recognises the importance of free movement of such data, and thus facilitates this through harmonisation. As a result, it generally allows for free movement of data, but regulates its processing and confers protective rights to the data subjects, e.g. the right to rectification (Article 16) or the right to be forgotten (Article 17).

Under the GDPR, personal data may be collected and processed for 'specified, explicit, and legitimate purposes' and 'not further processed in a manner that is incompatible with those purposes'. Whether or not a purpose is 'incompatible' must be assessed against several indicators (Article 6) – one of these being 'the existence of appropriate safeguards, which may include encryption or pseudonymisation'.³⁸

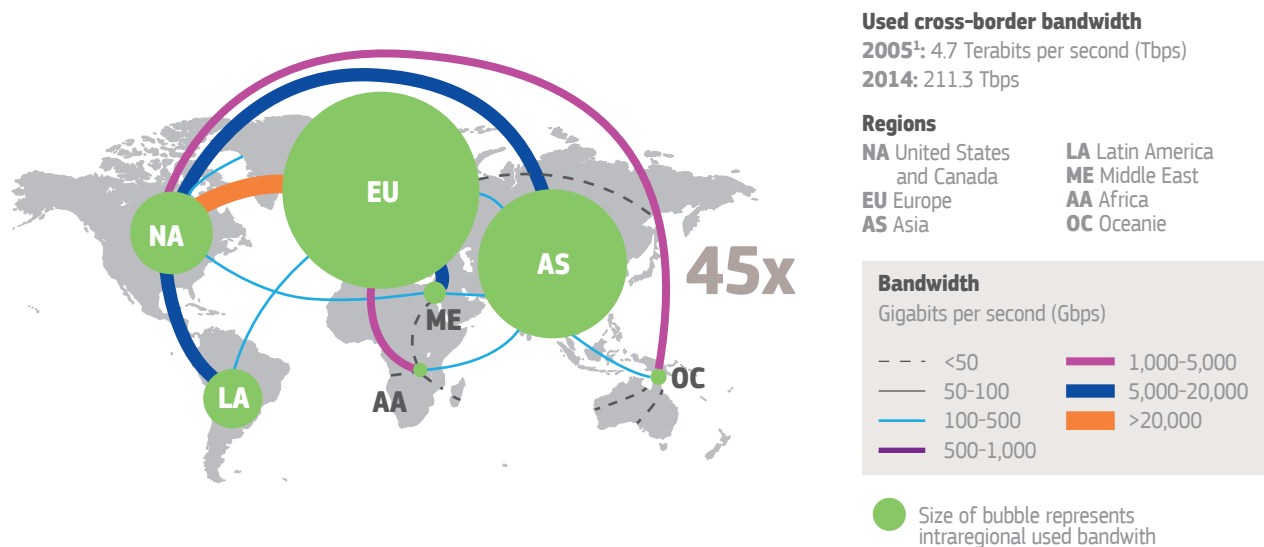
This new, Europe-wide data framework is long overdue and has the potential to yield significant benefits, providing, for the first time, a unified, common data regime for all of the Member States. **One cause of concern is the tendency by national governments to 'goldplate' European legislation, adding unnecessary complexity and increasing legal uncertainty. This should be avoided at all costs because in the digital age, speed and scale are decisive factors.** Goldplating of the GDPR at national level would reintroduce fragmentation and therefore slow down data flows, both of which would defeat the very purpose of the GDPR. Only a truly unified, agile and user-friendly legal framework can be the much-desired game changer for Europe's data economy.

Tackling restrictions to data flows

The GDPR creates a uniform level of data and privacy protection in Europe and, in a second step, prohibits restrictions affecting the free circulation of personal data within the Single Market based on these grounds. However, a part of data underlying big data analytics in the 'Internet of Things' markets are of non-personal nature and are therefore outside the scope of GDPR. Data can also be subject to localisation requirements that are not motivated by privacy protection considerations but rather based on public policy reasons (e.g. taxation).

A 2015 McKinsey report estimates that Europe is by far the region with the highest interregional data flows in the world (Figure 7).³⁹ Yet, Europe has been unable to fully unleash the potential of the data economy in large part because of **'digital protectionism'** in some Member States, i.e. when public authorities impose data localisation requirements on companies not for privacy protection or security concerns, but simply to limit cross-border competition. While this is ostensibly done to protect domestic businesses, in reality such actions are to the detriment of traditional industries (retail, manufacturing, energy), which stand to benefit greatly from digitalisation and cross-border data flows.

Figure 7: Europe has highest interregional data flows



Source: TeleGeography; McKinsey Global Institute analysis

Two thirds of respondents to a public consultation launched by the European Commission indicated that national restrictions to data flows and data localisation requirements have affected their business strategies. Forcing companies to store their data in the country where it was collected can increase costs by up to 120%, while differences in processing costs across Member States can also be significant. The Leviathan Security Group calculates that the cost of data processing in Germany is twice as high as that in neighbouring Belgium.

The removal of geographical barriers and the creation of a fifth freedom of movement for data within the single market should thus be at the top the Union's policy agenda.

From data ownership to open data

The allocation of property rights over machine-generated data to data generators (e.g. the owner of the car or the industrial user of a machine) presents a possible means of **giving users control of their data**. This could be seen as a means to facilitate trade and the creation of data markets. It would also seem to respond to a perceived need to ensure a fairer distribution of value across the data value chain, where currently data services and data users (in the car example – car manufacturers) are able to drive great profits from machine-generated data without sharing them with the data generators, despite the latter being indispensable actors of the chain. Such logic would be consistent with the approach observed elsewhere in the context of the implementation of the Commission's Digital Single Market strategy; for example, the proposed new neighbouring rights for publishers or rules on fair remuneration of authors and performers in the context of the copyright reform.⁴⁰

In the case of data however, **the creation of ownership rights would entail a number of risks.** Firstly, it may lower companies' incentives to invest in data analytics as their share of the value creation decreases. Secondly, it would face significant pragmatic challenges. The value of data is extremely dependent on the sample size: while all water drops are needed to generate the ocean, the marginal value of a single drop is negligible. Since the biggest part of the value is generated through aggregation, it would be difficult to craft a rule to 'fairly' assign part of that value to the single generators of the data. As Croll points out: 'the important question is not who owns the data' but 'who owns the means of analysis'.⁴¹ Furthermore, the more the data economy evolves, the more the bulk of the value will tend to be generated in the 'last mile' of the big data value chain.⁴² Finally, the introduction of data property rights would incur high implementation costs for single individuals and SMEs wishing to enforce them and would necessarily involve transaction costs that would reduce the incentive to trade the data. It would thus need to be seen whether and how such costs could be mitigated, for example, through standard contract terms.

Against this backdrop, an alternative route could be considered, in which the regulatory framework would – while respecting the framework of the GDPR regime – **maximise accessibility to data by any entity capable of generating value from it.** Initiatives such as 'open government', which see public administrations release data into the public domain respond to that logic. It should be considered whether open data regimes should be extended to the private domain, to the degree that allowing access does not undermine the investment made by the data services that first processed the dataset. As an illustration, consider mobility data harvested by telecoms companies. Anonymised aggregated samples could be shared with authorities or researchers to develop applications capable of monitoring and preventing the evolution of pandemic threats. Such experiments already took place in regions of Africa to predict where new outbreaks of Ebola cases were most likely to occur, based on the analysis of population movements stemming from mobile phone data. This helps to better plan aid interventions and respond to emergency situations.⁴³ Public intervention could prompt data-sharing, while also ensuring a limited scope for liability in the case of data-sharing for public interest. Access conditions would be designed such that the (often small) marginal costs for data sharing are covered, while incentives to invest in data collection are preserved.

Data portability and interoperability

Increasing competition at the data service level could transfer value to data generators and data users and thus emerges as a **more promising avenue to address issues concerning the fair distribution of value across the value chain than the definition of ownership rights.** For example, drivers may have restricted access to data harvested by satellite navigation devices as a deliberate choice of the device's manufacturer. However, increasing competition at the manufacturer level would naturally create an incentive to provide a wider access to data, since data access may become a parameter of competition.

Competition can be stimulated through the design of a regulatory framework that promotes *interoperability* of platforms, for example through open application programme interfaces (APIs). These interfaces allow applications to 'talk' to each other by making some of a programme's internal functions accessible to the outside world without requiring developers to share all of their software code (thanks to an API, for example, Yelp is able to display nearby restaurants on a Google Map).⁴⁴ Likewise, competition can be stimulated through *portability*, i.e. by ensuring that data is easily transferable from one service to another by their users. This means enhancing transparency in data processing and favouring the adoption of common standards, for example for data formatting so that different services can easily read or map the same dataset on their own platforms. These principles are already part of the regime governing personal data under the GDPR.⁴⁵

Data portability and interoperability are even more important in the field of machine-generated data, and accompanying **rules are urgently needed so as not to lower services' incentives to invest in data-driven innovation.** Limits should be envisaged whenever data transfers would entail undermining a service's business model. Take, for instance, the case of shopping platforms. Often sellers build a reputation thanks to a buyers' feedback system enacted by the platform for each performed transaction. The accuracy of the feedback system is one of the key parameters of success of the shopping platform. If sellers were able to export their ranking from that platform to a competing one, this would likely deplete that platform's initial investment: new players would have a lower incentive to develop new and better feedback systems if they know that this may also benefit their competitors.

Box 4: Personal Information Management Systems: A promising solution to data portability?

Personal Information Management Systems or PIMS consist of a user's server, running the services selected by the user, and storing and processing the user's data.⁴⁶ In most cases today, users hand out information to different data services and platforms that then use that information to run the relevant applications. For example, a user transfers data to Facebook and Facebook uses that information to provide social networking services. The user then supplies the same and other information to multiple platforms, creating a very fragmented system, often escaping the users' immediate control.

With PIMS, users would have a **personal digital deck** where all their information is stored. Services (such as Facebook) would then run on this deck, giving users the ability to keep track and control the information they share and, above all, easily use that information for multiple platforms. Hence, PIMS have the potential to significantly increase transparency and portability of data and, therefore stimulate data service competition.

A real-life experimentation with PIMS organised in France in 2013 with 300 users, resulted in the development of 15 prototype applications and the exchange of 5 million personal data sets.⁴⁷

A fair allocation of data liability

Rules on the allocation of liability play an important role in stimulating the development of data-driven business, particularly in the case of the emerging Internet of Things markets. Systems of interconnected devices tend to be highly integrated with each other and data – being the 'blood' flowing through the system – is the actual expression of that interdependence. A clear allocation of responsibility in case of damage (for example, for an erroneous use of data, wrong programming or flawed datasets) may help to inject certainty in markets where actors find it difficult to predict the risk entailed by their investment in data-driven activities.

As suggested by Bertolini,⁴⁸ it should be possible to separate safety issues from damage compensation issues. In the latter case, rules could follow a simple economic principle: **liability should be allocated to the party that is best placed to minimise costs and litigation**, to provide compensation and to ensure product safety. This would be the actor in the value chain – be that the user, the manufacturer of the machine or the programmer – that can most easily anticipate and manage the risk of failure in the data value creation process. In the case of autonomous cars for instance, the question is often raised whether the car manufacturers or the software providers should bear responsibility for potentially harmful malfunctioning of the car. Taking into account the fact that safety is better ensured by requiring manufacturers to conform to detailed technological standards defined *ex ante*, and that reputational risk acts as a major market incentive to safe design, it could be worth arguing that producers should bear absolute responsibility. This would also prompt them to enter into insurance schemes – the cost of which can be anticipated and factored in by the producer, while at the same time ensuring that users will receive compensation in case of damage.

Step 3: Active public policies to support the digital transition

As outlined above, the adoption of data technologies and data-informed decision-making is hindered by a number of barriers both within and outside organisations. The European Commission, Member States and regional/local public administrations can play an active role in removing these barriers.

In 2014, the European Commission Communication *'Towards a thriving data-driven economy'*⁴⁹ set out concrete actions to support and accelerate the emergence of the data economy. Among the measures proposed was the development of a 'big data' community, based on public-private partnerships, as well as 'excellence networks'; support for the deployment of necessary ICT infrastructure, including cloud computing; and an initiative to promote the adoption of open data and big data analysis within public administrations. The spirit of the 2014 Communication is well captured in the Commission's *Digital Single Market strategy*, section 4.1 – *building a data economy*.⁵⁰

The GDPR is an important milestone towards this goal but it needs to be complemented with policies that signal overall openness for the data revolution underway. To reap the full potential, a compelling vision is needed for what Europe's particular niche and competitive edge can be. It will certainly be on the high-end of the data value chain, building on the digital confidence that high data protection bestows, while also ensuring a fully operable, user-friendly foundation, spanning the entirety of the EU market. Operating at the interface of 'privacy by design' and the largest,

soon-to-be unified data market in the world could finally provide Europe the scale and distinctive feature that has hitherto been missing.

Understanding the data phenomenon by creating a better evidence base

Developing a data policy requires a good understanding of the evolution of the data economy through a systematic collection of statistical information on the use of data analytics by European companies. The European Commission should **ramp up the existing European Data Market Monitoring tool**⁵¹ by embedding it in Eurostat, the European statistical agency, and expanding its scope with new and more sophisticated indicators. That could include companies' detailed use of cloud services; the presence of data officers; number and severity of data incidents (such as data breaches); the use of big data analytics via machine-to-machine appliances; total spending on big data analytics; total amount of fixed investment in ICT infrastructures for data analytics; statistics relative to labour market dynamics (e.g. number of positions for employees with data-related expertise opened and/or fulfilled) etc. It would then be possible to monitor over time the evolution of data markets and provide a sound assessment of the effectiveness of public support policies.

Setting the right targets

Any data strategy should include well-defined medium and long-term targets, relating, for example, to the total value of data markets (e.g. they should reach at least 5% relative to GDP by 2025, against 1.87% in 2015), or to the adoption rate among the population of companies (e.g. the number of companies making intense and strategic use of data analytics should reach 30% by 2025, against a rate of 6.3% among ICT companies and professional service industries in 2015, according to the European Data Market Monitoring tool). What's more, **just as one of the key five targets of Europe 2020 is to ensure that at least 3% of the EU's GDP is invested in R&D, the EU could set clear and ambitious targets to stimulate companies' investment in digital technologies in general and data analytics in particular.** As pointed out by the OECD,⁵² data and R&D share a number of common features: both are intangible assets and non-rival goods that can benefit multiple users, but their production tends to be socially sub-optimal because the private incentives of companies are not aligned with those of society. It would hence seem only natural to deploy an overall data investment strategy as part of the EU's R&D strategy.

Promoting a mentality shift

Lack of awareness of the benefits of the data revolution remains a major barrier in companies, in particular SMEs. Companies simply do not know what is possible with big data.⁵³ Likewise, to embrace data analytics, companies must be able to win over internal resistance to change and implement a mentality shift – for example, different departments must be ready to openly share their information within the organisation (subject to the purpose limitations defined by the GDPR in case of personal data). Public authorities at EU and regional level can help through mentoring and educational programmes (for example through the deployment of dedicated web platforms) and through investment in far-reaching communication to promote success stories and raise awareness among the larger public. **Public administrations should also lead by example** by endorsing open data and big data principles in the administrative process, in line with established legislation and policy. Beyond what is already being done (for example in terms of 'open government data'), the European Commission should also show its commitment with concrete measures to reshape its internal organisation and to build up its own data analysis capacity.

One possible move could be to **create a European 'data analysis agency'** with high-performance computing capabilities to empower the Commission's Directorate-Generals with the technological means to process and analyse vast amounts of information that cannot be handed out to external services for confidentiality or cost-based reasons. This could be combined with the appointment of 'Chief Digital Officers' (CDOs) for all Directorate-Generals. CDOs would take responsibility for designing and fostering implementation of policies to digitise the administration. In particular, they would be entrusted with the task of ensuring the adoption of big data analytics within the organisation and to liaise and monitor the progress of national/regional/local administrations in that respect.

Developing the right skills sets

Another major barrier to big data uptake is the struggle to secure data-skilled staff. National education programmes should be adjusted to **increase the supply of data-scientists and a data-savvy workforce.** From an EU perspective, shortage of supply could be mitigated through measures aimed at facilitating the movement of data-skilled workforce through the single market, for example through the creation of a dedicated Europass for data-related skills. Students should also be given the opportunity to learn from the best educational and business centres across the continent: public-private co-funding mechanisms could be deployed to support students

willing to study at top European ICT universities or gain experiences in data centres. Efforts should also be made to **multiply ICT and data science courses in European universities**. The Union's Horizon 2020 research and innovation programme should be used to enable European research institutes and universities to make great advances in this field. Similarly to the recent establishment of the MIT Institute for 'Data, Systems and Society', funds should be allocated to projects aimed not only at developing new data-oriented technologies and analytics, but also at better understanding the interactions between the data revolution, systems-based theories and models, and the way in which societal processes work.

Financial support and incentives

Financing is another barrier that particularly affects SMEs when contemplating investments in data infrastructure and analytical tools, as these can carry high initial fixed costs. That challenge can be addressed by public administrations through appropriate funding, tax incentives and subsidies, although **European state aid rules would have to be updated accordingly**. For example, investments in research, development and innovation are recognised as an important driver of growth and a special regime is applied to state subsidies (up to 100% of the aid can be allowed depending on the total transferred amount, the purpose – whether experimental, industrial or fundamental research – and the size of the receiving company). Rules are based on the recognition *by default* that because of the inherent nature of R&D activities, subsidies are necessary to correct likely market failures. A similar special regime should be designed to support data analytics. Depending on companies' size, industry and level of internal development of data-driven decision-making, it should be possible to identify thresholds below which it is safe to assume that state financial support would have net beneficial effects – fostering data innovation to the general benefit of the economy, while having limited negative effects on competition.⁵⁴

Conclusion

The data revolution is already underway and gaining speed. It profoundly changes how value is created, with a relentless focus on user-centricity. At the same time, the very nature of innovation is changing, with data now a decisive factor in the success or failure not only of businesses, but also of the economies that underpin them. Going forward, Europe needs to extend the regulatory and legal certainty afforded to personal data via the GDPR to the fast-growing area of non-personal data. Above all, a comprehensive policy blueprint is needed on how Europe can accelerate its performance in the global data economy while finding its particular niche at the high-end of value creation.

Notes

1. European Commission, *'The EU Data Protection Reform and Big Data: Factsheet'*, March 2016.
2. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016/L 119/1, 27 April 2016.
3. The proportion of personal and non-personal information in business data is hard to quantify due, in particular, to the absence of a universally accepted definition for what belongs to these two categories of data. That appears to be an important question that would benefit in-depth academic research efforts.
4. Peña-López, I., *'World Development Report 2016: Digital Dividends'*, 2016 and Aaronson, S. A., *'The Digital Trade Imbalance and Its Implications for Internet Governance'*, Chatham House, 2016.
5. Ericsson, *'Ericsson Mobility Report: On the pulse of the networked society'*, June 2016, <https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf>.
6. Rosello, A., *'The Internet of Things, 3 Value Shifts Manufacturers Should Embrace'*, PTC blog, 7 March 2014, <http://blogs.ptc.com/2014/03/07/the-internet-of-things-3-value-shifts-manufacturers-should-embrace/>.
7. Deloitte University Press, *'The future of manufacturing Making things in a changing world'*, 31 March 2015, <http://dupress.deloitte.com/dup-us-en/industry/manufacturing/future-of-manufacturing-industry.html>.
8. Rosello, A., op. cit., 2014.
9. McKinsey & Co., *'Competing for the connected customer – perspectives on the opportunities created by car connectivity and automation'*, September 2015, <http://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-carmakers-can-compete-for-the-connected-consumer>.
10. Mandel, M., *'Data, Trade, and Growth'*, TPRC 412: The 41st Research Conference on Communication, Information and Internet Policy. The Progressive Policy Institute, March 2013.
11. Brynjolfsson, E., Hitt, L. M. and Kim, H. H., *'Strength in numbers: How does data-driven decision making affect firm performance?'*, SSRN 1819486, 22 April 2011.
12. European Commission, *'Fact Sheet Data cPPP'*, 2014.
13. European Commission, *'The EU Data Protection Reform and Big Data: Factsheet'*, March 2016.
14. From 2015 to 2020, the total benefit to the UK economy of big data analytics is expected to amount to £241 billion, or £40 billion on average per year, according to: SAS & Centre for Economics and Business Research Ltd, *'The Value of Big Data and the Internet of Things to the UK Economy'*, February 2016.
15. International Data Corporation (IDC) Research: *'Worldwide Big Data Technology and Services Forecast, 2015–2019'*, October 2015 forecasts that big-data related services will be worth \$48.6 billion in 2019.
16. Commission Staff Working Document SWD(2016) 187 final: *'Europe's Digital Progress Report 2016'*, 25 May 2016.
17. [European Data Marketing Monitoring Tool](#).
18. OECD, Science, Technology and Innovation Policy Note on *'Data-driven Innovation for Growth and Well-being'*, OECD Publishing, Paris, October 2015.
19. 93% of Chinese customers are willing to share their location data with the manufacturer of their car, compared to 65% of Germans and 72% of Americans according to: McKinsey&Company: *'Car data: paving the way to value-creating mobility - Perspectives on a new automotive business model'*, March 2016.
20. UK House of Commons Science and Technology Select Committee, *'Big Data Dilemma report'*, February 2016, and The Register, *'De-anonymising data should be a criminal offence, says MPs report'*, 12 February 2016.
21. Article 32 of Regulation (EU) 2016/679
22. When it comes to personal data, all of these actions will have to be taken in compliance with the GDPR, meaning that data protection principles will have to be adhered to.
23. Shapiro, C. and H.R. Varian, *'Information Rules A Strategic Guide to the Network Economy'*, Harvard Business Press, 1999.
24. For an excellent description of the transformational challenges faced by traditional industries see: Wessel, M., A. Levie and R. Siegel: *'The Problem with Legacy Ecosystems'*, Harvard Business Review, November 2016
25. OECD, op. cit. 2015.
26. TNO report: *'Thriving and surviving in a data-driven society'*, September 2013.
27. Papandropoulos, Penelope, *'How should price discrimination be dealt with by competition authorities?'*, Droit&économie Concurrences, N° 3-2007 – pp. 34-38.
28. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: *'A Digital Single Market Strategy for Europe'*, COM(2015)0192 final, 6 May 2015.
29. Independent Experts Advisory Group to the UN Secretary General: *'A World That Counts. Mobilising the Data Revolution for Sustainable Development'*, 2014.
30. Frischmann, B.M., *'Infrastructure: The Social Value of Shared Resources'*, Oxford University Press, 2013 and OECD, 2015.
31. Cf also special provisions on data concerning health in Article 9 GDPR.
32. For more information on this initiative please refer to the European Commission's Research and Development Information Service ([CORDIS](#)).
33. OECD, op. cit., 2015.
34. Rutkin, Aviva, *'Machine predicts heart attacks 4 hours before doctors'*, New Scientist, 6 August 2014.
35. DLA Piper Publications, *'Changes afoot in China cyber and data'*, 15 July 2016.
36. Regulation (EU) 2016/679.
37. The Regulation distinguishes between **different categories of personal data**. Those that are more sensitive include data 'revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health, or data concerning a natural person's sex life or sexual orientation'.
38. Pseudonymisation is 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.
39. Manyka, James, Mc Kinsey, Slide 12, speech delivered at 'Europe as Investment Destination', EPSC event, Brussels, 2016.
40. Commission [Proposal](#) for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, Brussels, COM(2016) 593 final, 14 September 2016.
41. Croll, A., *'Who owns your data?'*, Mashable, 12 January 2011.

42. TNO, op. cit., 2013.
43. See Wesolowski, A. et al, '[Containing the Ebola Outbreak – the Potential and Challenge of Mobile Network Data](#)', September 2014, as well as Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K. and Piarroux, R., '[Using mobile phone data to predict the spatial spread of cholera](#)', Scientific reports, 5, 2015.
44. Proffitt, Brian, '[What APIs Are And Why They're Important](#)', HACK, 19 September 2013.
45. Recital 68 of Regulation (EU) 2016/679 states that: 'Data controllers should be encouraged to develop interoperable formats that enable data portability. That right should apply where the data subject provided the personal data on the basis of his or her consent or the processing is necessary for the performance of a contract. It should not apply where processing is based on a legal ground other than consent or contract'. The data subject even enjoys a right to data portability under certain circumstances (Article 20), but it must not be forgotten that 'disclosure by transmission, dissemination or otherwise making available' of data is processing in the meaning of the GDPR and thus subject to the restrictions therein.
46. Abiteboul, S., André, B., Kaplan, D., '[Managing your digital life with a personal information management system](#)', Communications of the ACM, 58 (5), pp.32-35, 2015.
47. Fing, '[Self Data](#)', Cahier d'exploration MesInfos 2e édition, 2015.
48. Bertolini, A., '[Liability and Risk Management in Robotics](#)', Presentation to the European Parliament, 21 April 2016.
49. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: '[Towards a thriving data economy](#)', COM(2014)0442 final, 2 July 2014.
50. COM(2015)0192.
51. [European Data Marketing Monitoring Tool](#).
52. OECD, op. cit., 2015.
53. '[The biggest obstacle we're running into is not knowing what's possible. This is the single biggest problem we're running into,](#)' Praveen Kankariya, founder and CEO of Impetus Technologies, a developer of streaming big data analytic software and services based in Los Gatos, California, in: '[Knowing What's Possible a Big Obstacle for Big Data](#)', Datanami.com, 1 February 2016.
54. On this point see the Report '[The Role of Science, Technology and Innovation Policies to Foster the Implementation of the Sustainable Development Goals \(SDGs\)](#)', prepared for the DG RTD, 2015.