demos EUROPA
CENTRE FOR EUROPEAN STRATEGY

WARSAW INSTITUTE FOR ECONOMIC STUDIES

Big & open data in Europe

A growth engine or a missed opportunity?

Big and open data in Europe

A growth engine or a missed opportunity?

Sonia Buchholtz, Maciej Bukowski, Aleksander Śniegocki
the Warsaw Institute for Economic Studies (WISE Institute)

A report commissioned by demosEUROPA – Centre for European Strategy Foundation within the "Innovation and entrepreneurship" programme.

Macroeconomic modelling and publication sponsored by Microsoft.

Table of contents

Abstr

act

In this paper we discuss the economic potential of the current phase of the ICT revolution in Europe. We argue that a logical consequence of the exponential improvements in the computational power of integrated circuits and data storage capacities observed since the late 1960s, finds logical prolongation in the form of Big and Open data revolution. We support the view that ICT together with Big and Open data solutions reveal many features of General Purpose Technologies, and as such, have strong influence on the entire world economic system. We focus our quantitative analysis on Europe and with the help of a dedicated, bottom-up, macroeconomic model, which we call BOUDICA, we are able to estimate the economic potential of Big and Open data in all 28 EU member states. Moreover, we decompose the overall effect (estimated at 1.9% of EU-28 GDP by 2020) into 21 sectors and 28 countries showing that three parts of Europe – North, South and East – may benefit from the current technological trends in ICT differently. We argue that this divergent potential should be associated with different sectoral and structural bases of national economies in Europe, including different levels of innovativeness and vertical integration within particular sectors. At the end of the paper we thoroughly discuss policy implications of the major economic observations made in the article, pointing out the weaknesses and strengths of European industrial, competitive and science policy with respect to the Big and Open data challenge.

Execu

Summary

tive

The emergence of big and open data use is yet another phase of the ongoing ICT revolution which resembles previous great technology-driven economic transitions

Not all technologies are created equal. The majority of inventions are derived from previously implemented solutions, but some – the so called General Purpose Technologies – form a prerequisite for other inventions, imposing a substantial, long term impact on our wealth and quality of life. Such inventions should have potential uses in most branches of the economy, significant room for a cost cutting technological improvement and increasing availability, as well as be complementary to a wide range of other innovations.

Identifying truly transformational technologies the moment they emerge is a daunting challenge. Only after a dozen years or even several decades does the economic impact of new solutions become discernible. This was the case for electricity over a century ago and is apparent now for the information and communications technology. The emergence of big data enabled by the recent hardware and software advances and complemented by the shift towards openness provides yet another example of the ICT revolution persistence. While creating new opportunities for the European economy, this next phase of the ICT-driven economic transition also implies new developmental challenges for the EU as a whole as well as for its member states and their regions.

Big data solutions are moving away from solving the technical problems to shedding light on a new quality of insights

The concept of big data is usually defined by the "three Vs": volume, velocity and variety. In many areas volumes of available facts are higher than ever before, they are also expanding quicker, come from many more sources and materialize in many different forms than small, well-structured data sets from the past. With the growing use of big data in business, an additional "V" became important – veracity. In contrast with the three original Vs, it refers not to the intrinsic characteristics of big data itself but rather to the quality which makes it useful for practical applications.

The progress of computer technologies is constantly moving the frontier of what is *troublesomely* big, expanding the variety of *valuably* big data. The fundamental difference and source of the lasting value of big data is reaching the point where we get the *whole, live* and *expandable* picture of a given phenomenon. *Whole* data allows the capture of intricate aspects of the given process or study the subgroups of interest. *Live* data shows detailed unfolding of the process in real time and gives us a chance to immediately react to unexpected circumstances. Complexity and numerous linkages between different phenomena mean that data concerning one occurrence can provide important information about the other, so big data analysis is inherently *expendable*. Positive network effects occur: it is possible to gain fresh insights from constant building up, recombining and extending the scope of the analyzed data. Furthermore, all these improved insights can be used to make targeted predictions and decisions. The eye of the data scientist sees both a wider and a sharper picture than even before.

The business applications of big data are based on companies utilising new and improved insights. The resulting economic gains can be put into three broad categories:

- **resource efficiency improvements** through reducing the information concerning resource waste in production, distribution and marketing activities,
- **product and process improvements through innovation** based on R&D activities, day-to-day process monitoring and consumer feedback,
- **management improvements** through evidence-based, data-driven decision making.

From **Big Trouble to Big Value.**

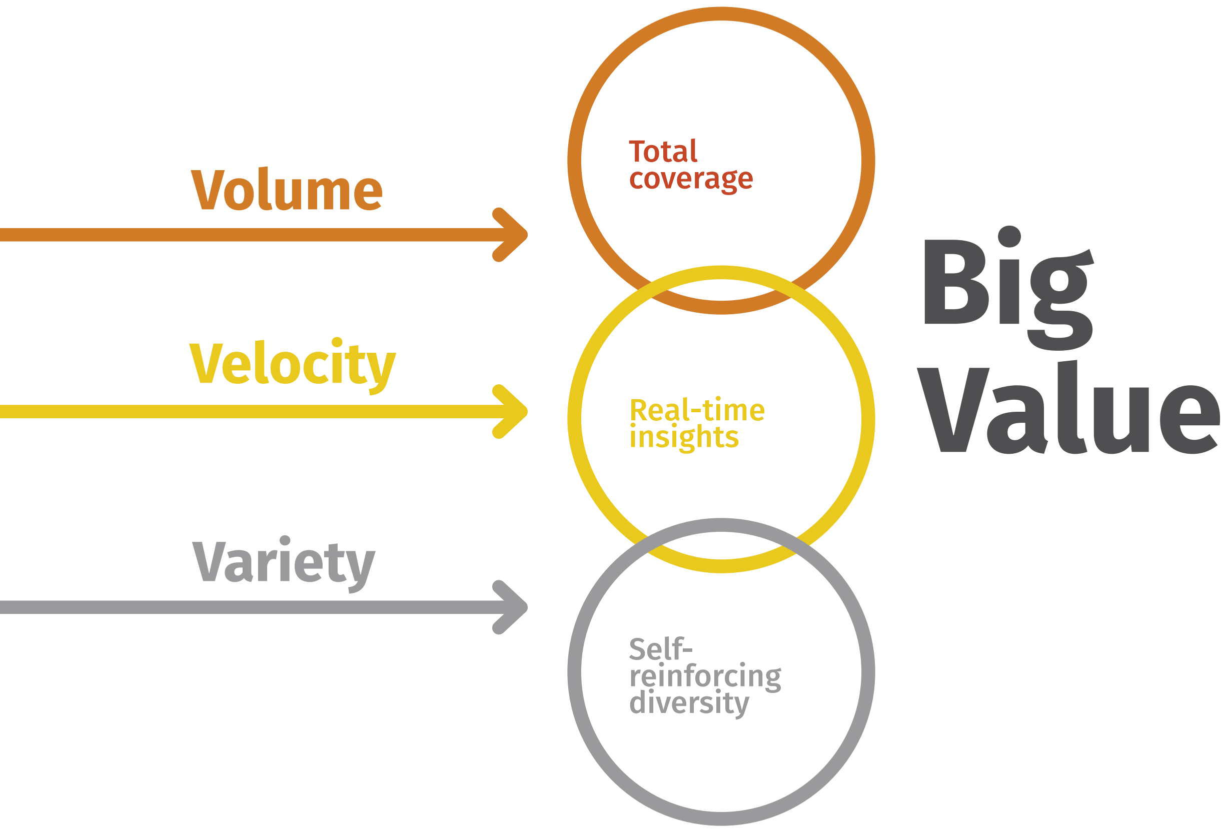
3 Vs transformation

Source: WISE Institute

The channels through which big data affects the economy are closely linked to the above mentioned categories. There are three ways in which data can turn into valuable input for a company:

- **Data-to-Information** is the key source of big data value for business models where information is the core source of value. The data is "mined" for valuable information and the value is created the moment the search succeeds.
- **Data-to-Product/Process** effect occurs when insights from data analysis need to be implemented in the physical world to bring value to an enterprise.
- **Data-to-Management** systematically brings data-based information into a company's decision-making process.

Data-to-Management is the most challenging way for companies to translate insights from big data to value as a company needs not only to successfully mine the data and adapt the insights for the management purposes, but it also needs to adjust its corporate culture and move from instinct – to data-driven decision making. This source of big data value is open for companies throughout the economy, though data-intensive sectors with high rates of ICT adoption may gain the most.

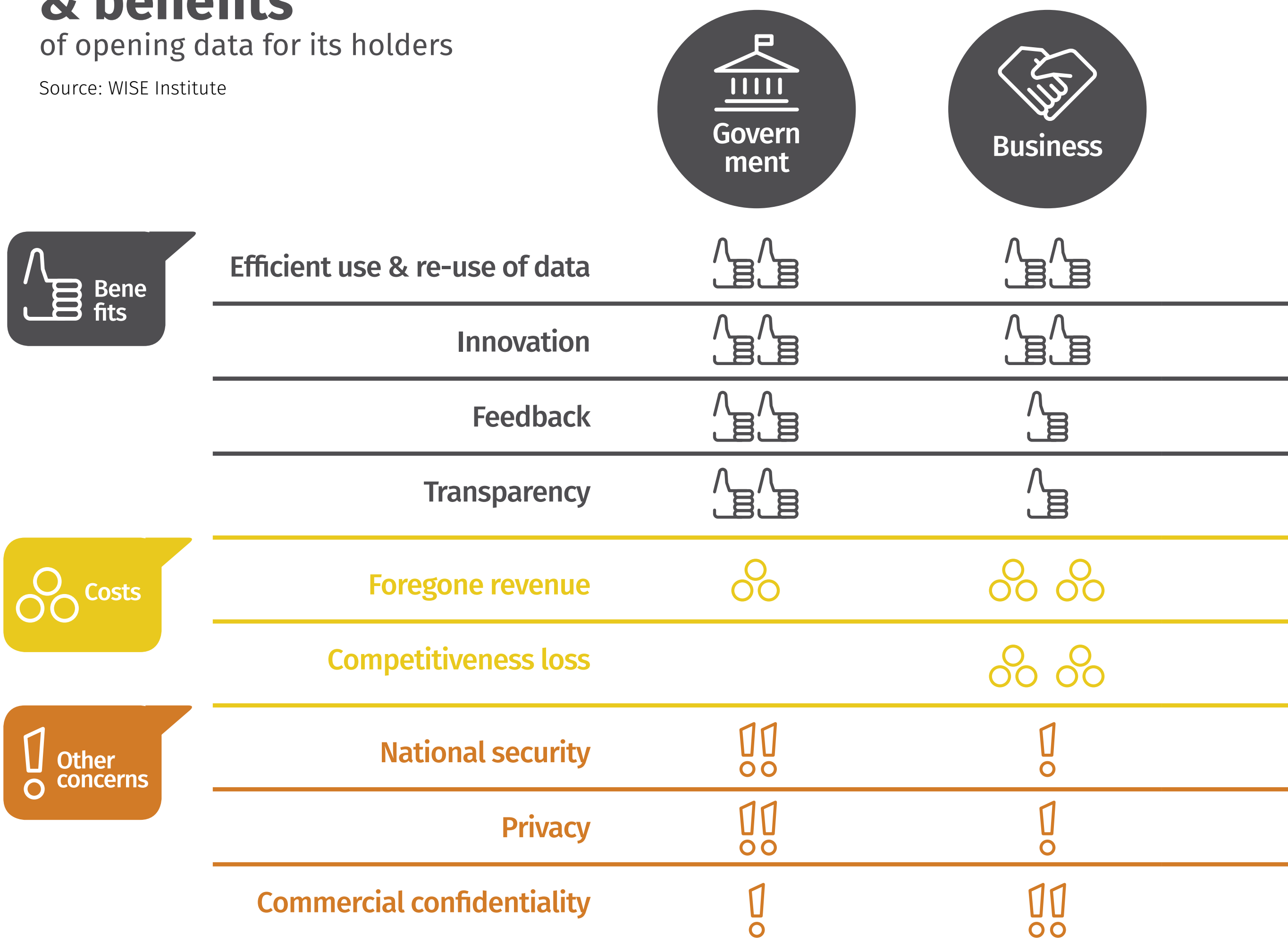


Openness of the data is complementary to the shift to big data solutions

Costs & benefits

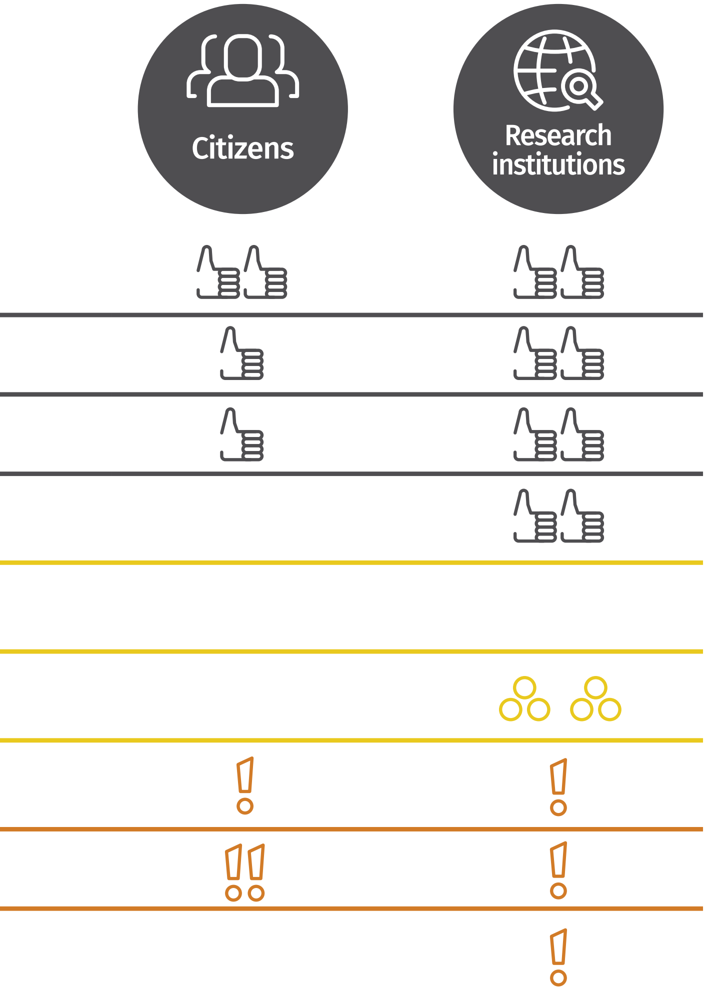
of opening data for its holders

Source: WISE Institute



Opening up and ensuring easy access to data for all parties provides several benefits both for data holders and data users:

- **Rediscovering the value and purpose of data** through use, reuse, linking and recombining it with other sources. Novel uses can be found long after the data was collected and utilized for its primary purpose.
- **Fostering innovation** by providing maximum ease of access to data cultivates experimenting and innovation. While data is closed, it is impossible to assess its full potential or mine it for unexpected correlations and insights.
- **Gaining feedback** by opening data up as a way to complement internal analytics with the insights from the external sources.
- **Increasing transparency** leading to strengthened public control and an enhanced debate on the government's actions, increased reliability of scientific results and a rise of businesses' credibility.
- **Gaining from the network effects** as more elements can be recombined and experimented with. Each new open piece of data adds value to the whole ecosystem.

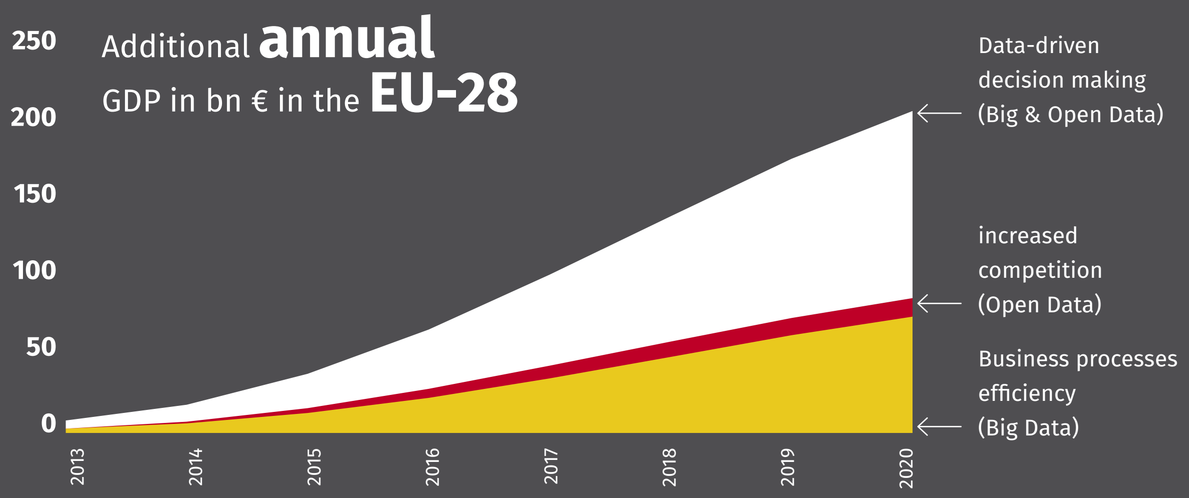


The key principle of open data philosophy is ***open by default***. Instead of finding a reason why the given piece of data should be open for the public to use and re-use, its owners should start with the default notion that data should be open unless there are important reasons to restrict access to it. Thus, there is a paradigm shift from what can be called *positive data freedom* (defining explicitly what can be opened) to *negative data freedom* (defining explicitly what should be closed). Beside the direct cost-benefit considerations, there are three main reasons *not* to open data. These are protecting citizens' privacy, not violating commercial confidentiality and addressing national security concerns.

The successful shift to the negative data freedom paradigm requires clear universal rules for when *not* to open data and which ways of using it should be prohibited. Without them, there will be only islands of openness in the ocean of closed data.

EU-28 Additional GDP by 2020

206 bn €



Source: WISE Institute with BOUDICA

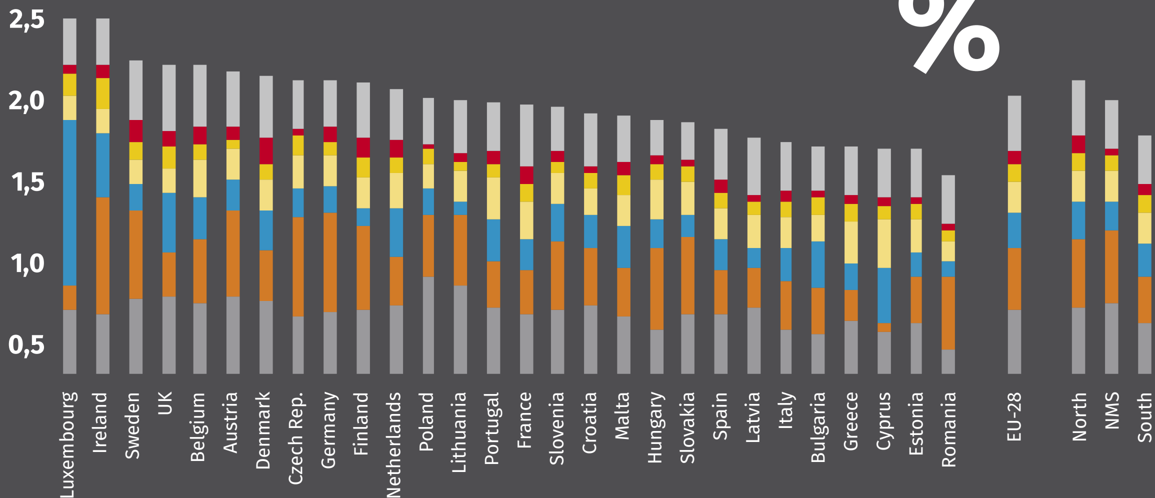
Additional **output** in 2020

	Total bn €		per capita €
	150	North	560
	16	NMS	148
	40	South	311
	206	EU-28	408

Trade
Manufacturing
Health
Information & communication
Finance & insurance
Public administration
Other

Additional **increase** in GDP level by country and sectors

%



The EU economy will grow by an additional 1.9 % by 2020 thanks to big and open data

focus on 3 key forces

In order to estimate big and open data impact on the EU, we focus on three key forces which affect the economy on the aggregate level and incorporate them into the bottom-up BOUDICA model. These effects are:

productivity increase in manufacturing & services

resulting from increased business processes efficiency

competition improvements

resulting from lower entry barriers for business as a consequence of the opening of public sector data,

improved allocation of production factors

resulting from better decision making thanks to shifting from instinct – to data-driven management processes.

Model results indicate that by 2020 big and open data can improve European GDP by 1.9%, an equivalent of one full year of economic growth in the EU. The majority of this effect will depend on big data applications in management and resource efficiency. The direct macroeconomic effect of open data will be relatively modest, but it will play an important role as complimentary element of the big data revolution.

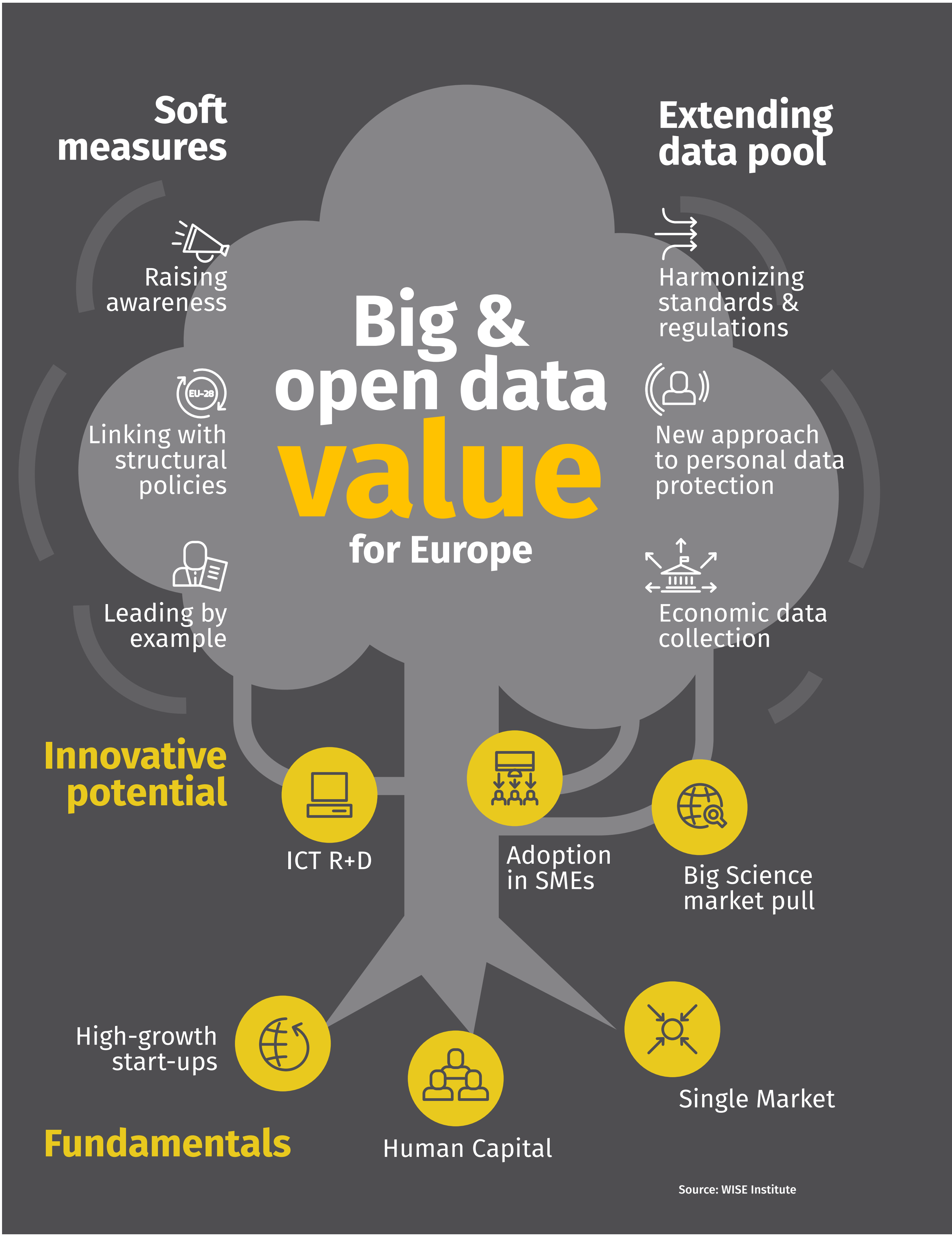
Countries with larger, more global enterprises as well as those more advanced in the ICT field should benefit more than those that already lag behind in these areas. Thus the positive impact of the data revolution will be felt more acutely in Northern Europe – except France – while most of the New Member States and Southern European economies will benefit significantly less, with two notable exceptions being the Czech Republic and Poland.

The broad mix of policy measures should be implemented both at the level of the EU and individual Member States to ensure that Europe will unlock the full potential of big and open data

Triple challenge

1. recognize the different initial readiness
2. policy choices tailored to the local specificities
3. big and open data challenge as part of broader reform agenda

Europeans face the triple challenge with respect to big and open data. Firstly, they must recognize the different initial readiness of individual Member States to the challenges and opportunities of a data-driven economy. Secondly, they should address them with policy choices tailored to the local specificities, but at the same time strengthening the scale of the common market. Thirdly, they must look at the big and open data challenge not as an isolated regulatory challenge but rather as part of broader reform agenda for Europe that should reinvent our thinking about the post-crisis economic reality, and provide new engines for economic growth for the future.


Soft measures
Raising awareness
Linking with structural policies
EU-28
Leading by example
Extending data pool
Harmonizing standards & regulations
New approach to personal data protection
Economic data collection
Big & open data value for Europe
Innovative potential
ICT R+D
Adoption in SMEs
Big Science market pull
Fundamentals
High-growth start-ups
Human Capital
Single Market
Source: WISE Institute

Key areas, growth opportunities & challenges
for big and open data in Europe

Key areas

Switch to data-driven decision making

Ensuring that SMEs will have access to big data solutions (e.g. by third parties)

Efficiency improvements in manufacturing and trade

Growth opportunities

Re-igniting industrial growth through data-based innovative solutions

Improving public sector efficiency, fostering economy restructuring

Fostering knowledge-intensive services growth

Challenges

Securing talent for global competition

Achieving scale effects, securing financing for innovative data solutions, closing the ICT gap

Decreasing innovation lag, enacting effective open data policies

Source: WISE Institute

Introd

Recent years have witnessed an emergence of the big and open data debate.

What used to be an academic topic with narrow practical applications now appears to be one of the key transformational economic forces of the XXI century. Big and open data should nolonger be associated solely with the Internet, as its recognition stems from rather successful applications in non-ICT companies (improvement of inventory management and production processes through extensive usage of sensors), scientific research (processing vast datasets in physics, genetics or astronomy) or even politics (harnessing data on voters'behaviour to build effective campaigns). Wherever a computer algorithm completes a phrase in a search engine or recommends a book in an online bookstore, big data solutions are involved. The same can be said about the continuous monitoring of production processes or about the real time tracking of the economic forces that shape the economy onthe macro scale. Data helped to win the last presidential elections in the U.S., to find the Higgs Boson in Europe, and to design a computer algorithm capable of successfully competing in the Jeopardy contest against humans. Many claim that we are on the verge of a revolution that will completely transform our lives in the near future and that big and open data hide a key to its deployment.

In this paper we discuss the current development stage of the digital revolution in Europe,

arguing that ICT together with big and open data solutions reveal many similar features to the seminal General Purpose Technologies of the past such as steam power or electricity, and thus should have a large impact on the future prosperity of the world. The major goal of our

uction

The article is constructed as follows.

analysis was to provide quantitative analysis of the European ability to exploit the economic potential of big and open data within the next decade. To this end we have constructed the bottom-up macroeconomic BOUDICA model that allowed us to estimate the probable economic effect of the transition to data-driven solutions in the EU member states. We managed to decompose the overall effect into 21 sectors showing that three parts of Europe – North, South and East – are probably differently equipped to benefit from current technological trends in the ICT industry. We argue that this divergent potential should be associated with a different sectoral and structural background of national economies in Europe, including different levels of innovativeness and vertical integration within particular sectors. At the end of a paper we thoroughly discuss policy implications of our observations, pointing out the weaknesses and strengths of European industrial, competitive and science policy with respect to the big and open data challenge.

1

In the first chapter we describe historical examples of General Purpose Technologies looking for the analogies between their deployment in the mid-19th and early 20th centuries and the currently unfolding digital revolution.

2/3

In the second and third chapter we argue that big data constitute the next, natural stage of ICT development, and that open data plays an important complimentary role in this process. In both cases we discuss the major economic forces involved in the propagation of digital inventions into real world businesses.

4

In the fifth chapter we incorporate the representation of these mechanisms in a formal economic model that we subsequently use for the assessment of the potential of big and open data in Europe.

Summary

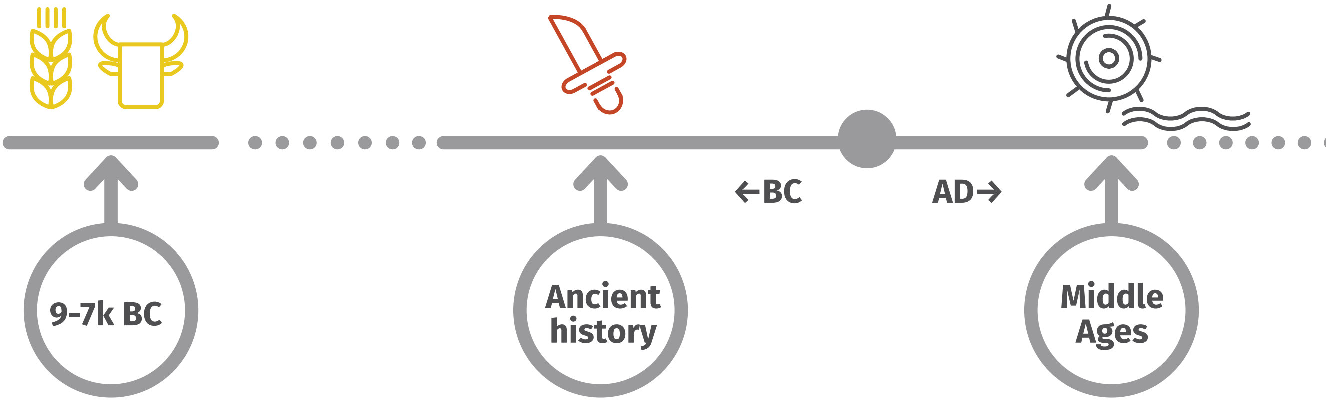
Last chapter includes a summary of our findings and presents policy recommendations.

1

Technological revolutions and

general purpose technologies

Revolutionary inventions considered as GPTs

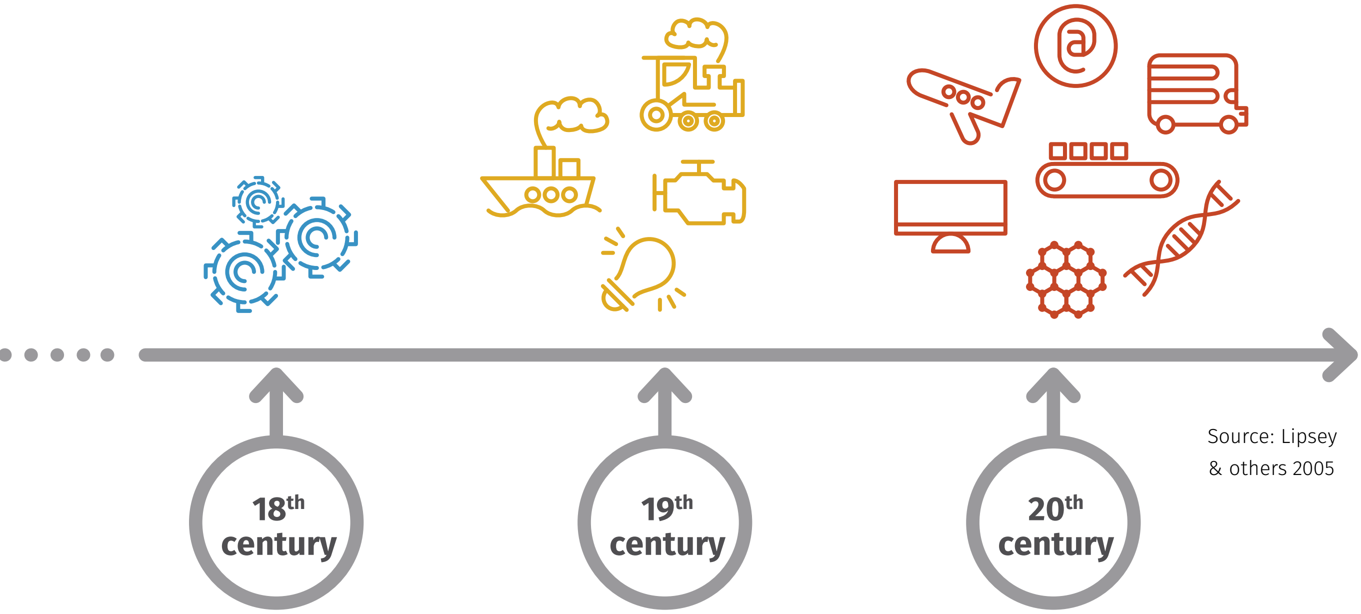


GPTs vs. other technologies

Nowadays we are constantly flooded with news and reports on technological inventions. Some of them seem to be qualitatively new, whereas others only extend and modify existing solutions. From the policy maker's perspective, only those innovations are of particular importance that generate strong incentives in the economic or social sphere, change companies' competitiveness and their ability to generate profits, or shift consumption patterns and political views of average citizens. The majority of these inventions are derived from previously implemented solutions, but some – the so called General Purpose Technologies (hereinafter GPTs) – form a prerequisite for other inventions, imposing a substantial, long term, impact on our wealth and quality of life.

First and foremost GPTs open the doors for other – less universal – inventions. This feature results from their origin. At first, inventions seem to be crude, incompatible with existing capital stock, and of low usefulness. As time goes by, improvements facilitate their implementation and commercialization in a broader range of branches thereby reducing unit costs and further increasing their pace of spreading. Furthermore, the increased range of use increases the probability of subsequent innovations. GPT seems to be an evolutionary experience. To put it in a different way, GPTs can support implementation of other inventions that, developed and commercialized, are changing the world. An especially large potential seems to lie in creating opportunities rather than in delivering finished goods (Bresnahan and Trajtenberg 1995). In this sense there is a similarity between GPTs and basic science (Clarke et al. (2006)).

This description delivers intuitive criteria for identifying GPTs. The most often cited definition of Bresnahan and Trajtenberg (1996) provides three of them. Invention: should be (1) **pervasive** – used in most branches of the economy, and have (2) **potential for technological improvement** – resulting in cost reductions as basic technology becomes developed and its availability increases, as well as (3) reflect **complementarity towards other innovations**



Source: Lipsey & others 2005

– enabling, creating and producing new products and processes. Distinguishing GPTs from other technologies stems from the extent of their use, the evolutionary approach towards their developments and ability to generate spillover effects. Similar conditions were proposed by Lipsey et al. (2005). Bresnahan and Trajtenberg (1995) emphasize technological dynamism, which is inherent in technical improvements, whereas Clarke and others (2006) highlight the issue of complementarity of innovations. Regardless of the exact definition, these criteria paint a quite coherent impression of what GPTs are, and it is easy to imagine that there have not been many inventions that fulfil all of them.

Lipsey and others (2005) classify ca. 20 candidates. The first one – the domestication of plants and animals –is dated to 9-7k BC. Antiquity brought us bronze and iron smelting, Middle Ages – the waterwheel. After the 17th century the increasing pace of life is reflected in a number of seminal inventions: at the turn of 18th century – 1, before the end of the 19th – 4 more and in the 20th century – 7 more. Analyses concerning GPTs often mention rotary motion (Bresnahan and Trajtenberg 1992), electricity and electric engine (Trajtenberg and Yitzhaki 1982), chemical engineering (Rosenberg 1998), electric dynamo (David and Wright 1999, 2005; David 2010) or binary logic and computers (Bresnahan and Trajtenberg 1992). It is worth noting that the criteria for enabling technologies are naturally met by material processing techniques (such as bronze or iron smelting), energy generation (steam power, electricity), tools (waterwheel), communication (computer, Internet), transport (three-masted sailing ship, railway) and organization technologies (mechanization). The following belong to the likely contemporary candidates: nuclear power, superconductivity, hydrogen fuel cells, robotization, biotechnology and nanotechnology. Nevertheless, not every promising invention deserves the name: lasers – although revolutionary in many applications, have not found mass use in the economy (Lipsey et al. 2005) and not one of the GPTs listed includes them. The case of GPTs shows that evolutionary development can generate a critical mass and a wave of spillovers. Quite naturally, revolutionary changes are rare, risky and often much more resource-intensive than minor modifications.

The past revisited[1]

Both revolutionary inventions and slight modifications are crucial in stimulating economic growth and welfare. Admittedly, technological change is only one aspect of a source of a long-term upwards trend of labour productivity. This observation may be proved by the thought experiment presented in Lipsey et al. (2005): If the size of the market and investments were fixed at the 1900s level and technological change proceeded as it has done for 110 years, we would be close to the contemporary standard of living. However, if we had allowed the market size and investments to change historically controlling at the same time the level and pace of technological advancements, our lifestyle would be much different.

The magnitude of GPT's influence may be approximated by changes they were responsible for – this, of course, applies only to the technologies whose utilization we can consider as complete. Admittedly, there are many such cases, but let us take a look at two: (1) discovery of the steam engine, which enabled mechanization of production and (2) discovery of electricity, which led to Fordism – one of the most important innovations in management. These cases can highlight the economic significance of GPTs, and provide us with a useful approximation of the potential for big and open data revolution in ICT.

The incentive to develop the steam engine was quite simple: the previously used sources of power were not able to satisfy a rapidly increasing demand. Decreasing profitability was a signal to search for more efficient energy sources (Carlsson 2004). Although the discovery of the steam engine was an effect of a private invention rather than public innovation policy and its first applications were limited to the mining sector, as time went by and first designs had been improved on and disseminated, a series of substantial shifts in the manufacturing management process took place.

First of all, increasing requirements for fixed capital in plants created an incentive for vertical integration in the industry. Moreover, independence from water energy sources reduced obstacles to plant location, facilitating economic development in many, previously rural, areas and leading to the structural shift from an agricultural to an industrial economy in just a few decades. Mechanization was also the reason for changes in labour demand, as only a limited proportion of workers had to acquire higher competences, whereas the majority could work efficiently without any schooling. Last but not least, the use of steam engines in transport and home appliances had its obvious contributions to the quality of life– for instance, increasing the consumption level and improving its structure or enabling longer leisure hours per day or week.On the other hand new working conditions shed light on matters of quality of life and child labour changing the approach to social matters in politics.

---

[1] subchapter largely based on Lipsey et al. (2005)

An even more obvious example of GPT is the electrical dynamo, quoted by Lipsey et al. (2005) and widely analyzed by David (David and Wright 1999, 2005; David 2010). Although its discovery in 1867 has not led to an immediate commercialization, a couple of decades later, electricity was present in almost all spheres of human life. Electricity enabled construction and installation of little energy sources which initiated a fundamental shift in the production process: the division of labour into small stages and ordering them along the assembly line. This step, pioneered by combined with well-defined tasks and senior managers' competences. Longer working hours with artificial light were also an instrument of human resources management. Both companies and citizens drew benefits from the increasing density of electricity. The diffusion of electricity seemed to be the main source of post-war economic boom in the USA. According to the analyses of David (1991), the main economic effects of this invention include boosted productivity, lower use of energy per unit of output and smaller consumption of energy per capita.

Electricity enabled an abundance of goods for the masses achievable for the first time.

Henry Ford in his automotive plant, reduced the role of manufactures and enabled mass production leading to the drastic reduction of unit costs of production in all sectors. For the first time an abundance of goods for the masses seemed achievable.

Electricity sources were much better suited for mass production than steam engines. The optimal volume of plants changed, fumes were reduced and working conditions improved rapidly. Oligopolies were favoured due to them reaping the rewards of economies of scale, the role of marketing rose whereas the production process had to be rearranged as a moving assembly line required implementation of new organization and management techniques,

ICT as a GPT

ICT seems to be a natural 20th and 21st century candidate for GPT

Most authors share this opinion, adhering to above mentioned definitions and historical examples of steam engine and electricity[2].

At first, electronic appliances were huge, cumbersome and demanding (Hempell 2006) but, as time went by and technological progress brought miniaturization and more widespread use, its carriers (i.e. computers) started to gain in speed, emerge in many branches and adapt to surroundings. ICT's broad influence on our economy should not be underestimated as its presence forced fundamental changes in production processes and opened many new business opportunities, through the exploration of new economic fields.

When it comes to (1) pervasiveness, not only is there a wide scope of branches where computers are used, but also a variety of uses (Lipsey et al. 1998), although it is emphasized that a broad range of applications does not equal full diffusion, which is far from being complete (Kretschmer 2012). On the other hand, as Jovanovic and Rousseau (2005) show, the pace of adopting is rapid and disparities are not unique for digital technologies. Analyzing (2) potential for technological improvement one may observe it in learning-by-doing characteristic for ICT technologies.

The most prominent example is boosting the computing power of microprocessors, which contributed vastly to the variety of applications of computers in the contemporary world (Hempell 2006). Taking look at (3) innovation complementarity – there are many inventions which were implemented thanks to the existence of ICT. In this context ICT enhances R&D productivity in downstream sectors, reaching a momentum where not only hardware complementarity is involved but also methods and workplace organization (Milgrom et al. 1991): economies of scale, marketing development, network, new role for management. Empirical evidence is ambiguous, though. Although the network effects of IT are straightforward in telecommunications, for ICT as a whole they may not be as obvious (Kretschmer 2012). On the other hand Oz (2005) points out that regardless of GPT properties, the 21st century companies simply would not survive without ICT investments.

Similarly to electricity in the late 19th century, no immediate effect of ICT could be observed in the macroeconomic data as of the end of 1980s. This observation led to the so called *computer paradox*, as researchers tried to prove that computers, although intuitively helpful at work, do not impact productivity data on the macro level (as Solow famously noted *'You can see the computer age everywhere but in the productivity statistics'*). This raised the question whether ICT based development is really a good choice for the economy (i.e. Berndt and Morrison 1995, Brynjolfsson 1994). The inflow of 1990s data falsified this approach (i.e. Bloom et al. 2010 – broad overview was prepared by Kretschmer 2012). Since then we are certain that ICT does have a significant, positive impact on the economy, however it took over two decades to see it materialize in the macroeconomic datasets.

If the impact is significant, but postponed in time, then an explanation of the lag between the moment of invention and its results is still required. GPT explanation assumes two conditions (Hempell

[2] There is also a group of antagonists of the ICT-GPT approach: Jorgenson and Stiroh (1999) found no spillovers (due to rapid deployment, as was added in Jorgenson et al. 2008), and Gordon (2000) found ICT unproductive due to the absence of spillover effects outside the durable goods manufacturing sector. In his later work (2003) this hypothesis is a bit weakened, when deducing exaggerated investments in IT.)

2006): that ICT is an enabling technology and that ICT requires a wide range of investment in intangible assets (institution structure, strategy, skills etc.). If these conditions are met, rewards from ICT investments should materialize over a longer period of time as the accumulation of factors necessary for full scale deployment of ICT productivity potential is gradual in nature. Structural changes, including restructuring and closer cooperation between companies, process optimization through new work contents or acquiring ICT skills simply take time (Autor et al. 2003; Spitz-Oener 2003; Chesbrough and Teece 2002). Moreover, investments in ICT as all investments in intangible assets cost money and resources whereas their accumulation is gradual. Therefore the initial impact of ICT investments on global productivity could have even been negative.

When taking a closer look at these intangible assets, companies' strategies concerning investments and ICT application, acquiring skills etc. seem to be an element determining the differences in how successful the implementation of ICT is. It is justified by the capabilities to use cheaper and powerful ICT in order to enhance productivity (Hempell 2006). Because of that ICT prefers highly skilled employees, contributing to the so called skill-biased technological change (Chennels and van Reenen 1999). On the other hand, the higher propensity to invest in ICT is correlated with human capital investments like trainings, cross-training or screen for education (Bresnahan et al. 2002). Finally, it is easy to suppose that the application of GPTs assumes a willingness to take risks. As a consequence, productivity (as conventionally measured) seems to be underestimated at initial stages of the process and overestimated in the latter, when ICT is fully implemented (Gordon 2003).

Macro-effects of ICT revolution[3]

From the macroeconomic point of view, GPT seems to create two channels fundamental for growth: (1) moving the technological or even the invention-possibility frontier, and (2) enabling some productivity gains, mostly by managing owned resources in a more efficient manner. Boosting each of these channels brings economic growth and, therefore, wealth and a higher quality of life and is reflected in larger output, labour productivity and total-factor productivity. Labour productivity is easy to measure, but does not adequately reflect all important details. On the other hand, total-factor productivity is prone to mismeasurement, omits intangibilities, as well as the time lag when externalities are materializing. Another problem is exacerbated endogeneity (as Kretschmar 2012 shows, evidence is mixed). All in all, enhancing either labour productivity or total-factor productivity should lead to higher output levels.

3 If we focus on ICT, the growth framework concentrates on **channels** through which **ICT can influence growth** (Hempell 2006):

1 by **technological progress** in the ICT sector: quality adjustment, input-excelled – increasing at lower rates (substantial productivity gains), As far as labour productivity is concerned, there are a few interesting analyses which state that differences in ICT saturation between US and EU may be the crucial factor accounting for differences in the GDP level. Timmer et al. (2003) call it the *Atlantic divide* and treats IT as the only factor influencing this difference, Daveri (2002) delivers estimations for new technologies. Other results are similar and the arguments put forward are: lower contribution in market services (retail and finance) (van Ark et al. 2008), as well as lower levels of ICT innovations (Inklaar 2008). Analyzing intangible capital controlling changes results, which means that IT does influence the labour productivity (Marrano et al. 2009, Oliner et al. 2007).

Special attention is devoted to spillover effects, as indirect contributions seem to have a bigger role than direct ones. It applies especially to network externalities (Brynjolfsson and Kemerer, 1996) and co-inventions. Bresnahan et al. (2002) suggests that the role of activities such as customized programming, training, inventing a business purpose and organization is so important that business co-inventions outstrip technical co-inventions and technical inventions. Another source of spillovers is the organizational complementarity. In this area many important conclusions have been worked out. First, there is a positive correlation between ICT and (1) service/output mix, (2) decentralized decision-making process and (3) well-skilled staff (Bresnahan et al. 2002; proportion of non-managers using computers – Black, Lynch 2001) and workplace organization (Black, Lynch 2001, Crespi et al. 2007). ICT has also been shown to reduce coordination costs, as well as the number of supervisors and to reduce capacity needs, although cost reduction seems to play a tangential role. Profits come from facilitated communication and new knowledge rather than reduced costs. That is compatible with a traditional opinion that ICT enhances productivity: R&D, education and knowledge spillovers etc., as mentioned i.e. in Romer (1990), Aghion and Howitt (1998). Moreover, services also seem to be a vehicle for spillovers (although manufacturing reaps such rewards; also Bosworth's and Triplett's (2003)), whereas strict regulations seem to have a negative long-term impact.

Stiroh 2002 states that higher intensity of ICT influences acceleration and labour productivity. If ICT investment correlates with subsequent product acceleration, it leads to productivity growth. Baily and Lawrence (2001) found this result outside the ICT sector (which equals to capital deepening and TFP growth). For single branches the results of digitalization and the Internet may differ, among them: product extension in traditional industries, restructuring, creation of more effective markets or new products and new branches. Cases of oil industry and banking presented by Carlsson (2004) are quite eye-opening, proving that there still exist space to use all the potential that ICT creates.

[3] subchapter largely based on Lipsey et al. (2005)

2

by **capital deepening:** transfers to downstream sectors, generating pecuniary externalities (Griliches 1992); this category refers to labour productivity only,

3

by **spillover effects:** non-pecuniary spillover effects in non-ICT-sectors (Stiroh 2002); innovative organization, processes, products, networking, co-inventions contributing to multi factor productivity.

2 Big Data

will it unleash the full potential of computers?

What is Big Data?

The concept of **big data** 3Vs

is usually defined by the volume, velocity & variety.

In many areas volumes of available facts are higher than everbefore, they are also expanding quicker than ever, come from many more sources and materialize in many different forms than small, well-structured datasets from the past. Therefore:

V **Volume**

V **Velocity**

V **Variety**

refers to large amounts of collected & processed data available for analysis.

It is important to note that this measure is highly context-dependent. Thanks to technological advances, datasets which were considered large only a decade ago today can be used on a regular basis. Therefore when thinking about the volume, it is useful to measure it in relative terms. What is qualitatively new about big data is that it covers the whole population (e.g. all clicks on the web ad, all transactions in the store) and not only a small subset of it. It allows to move from sample analysis to a more complete and detailed study of a given problem, as it is possible to detect much more complex and subtle patterns in the bigger, all-encompassing dataset.

refers to a high rate of data collection.

This makes real-time data analysis possible, and allows big data users to constantly refresh and improve their datasets. It is therefore possible to observe in great detail how complex processes gradually unfold. Wider and at the same time more granular data coverage not only in space, but also in time allows for refining the analysis. New short-term patterns can be detected which were previously obscured by a too low frequency of data collection. It is often possible to observe the results of a particular action or decision immediately (e.g. online stores) or much quicker than by using ordinary data techniques (e.g. monitoring the financial industry data flows to assess economic policy impacts). However processing high velocity data requires not only novel hardware and software solutions, but also high quality network connections which reduce latency.

refers to the numerous formats & sources of data as well as the fact that they are usually unstructured.

This is a challenge and an opportunity at same time. On the one hand, it is much more difficult to integrate data which do not come in a uniform format and structure. On the other hand, connecting diverse datasets forms a potential to uncover unknown details in the area of interest and shed new light on it.

V

The growing use of big data in business, an additional “V”, became important – **veracity**

In contrast with the three original Vs, it refers not to the intrinsic characteristics of big data itself but rather to *the quality* which makes it useful for practical application. There can be data which meet the three Vs criteria, but are inapplicable in practice due to poor quality, incomprehensiveness or low credibility of sources. At the same time the three Vs make the fourth – veracity – easier to achieve. Ordinary measurement errors are less of a problem when there are a number of independent measurements. Errors tend to cancel out and statistical techniques may be used to increase data robustness. Therefore the key to achieving veracity of big data is not ensuring perfect measurement, but rather avoiding systematic errors and controlling for trustworthiness of sources.

V

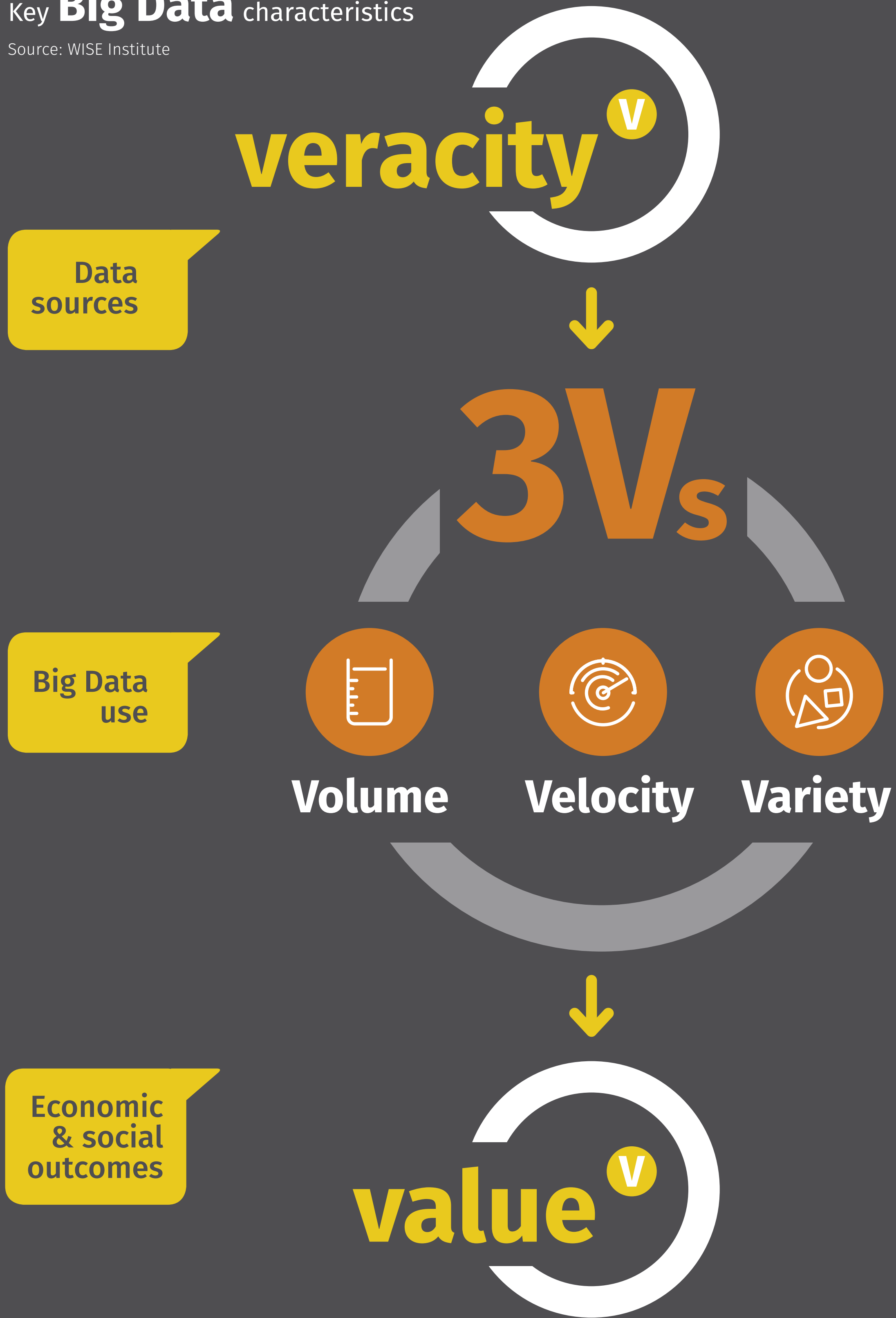
From the policymakers' perspective the key factor is another “V” – **value**

(OECD 2013)

It not only represents economic but also social outcomes of the widespread development of big data in the economy and the society. Exploiting the insights gathered from vast amounts of data can not only increase companies' revenues, lower operating costs or create new services, but also improve national security, decrease health risks and help policy makers to target their policies in a better way. The flip side are the privacy concerns which are generally greater in the areas where Big Data might be used for non-business purposes. There already is a heated debate about finding the right balance between increased security and scope of citizens' invigilation. Another important discussion about targeted data-based policy interventions is under way as it becomes technologically possible to select individuals or neighbourhoods where public intervention (social transfers, public investments) will probably be the most efficient. The need to optimize public expenditure will need to be balanced with concerns about equity and non-discriminatory state approach to the citizens.

Key **Big Data** characteristics

Source: WISE Institute



Why Big is Different – from technical problems to new quality of insights

The term “big” initially was associated mostly with data characterized by large volume. The anxiety concerning data deluge is still one of the main topics of the debate. Businesses, governments and scientists look at the growing mountains of data as a serious problem for traditional technological and organizational frameworks. The risk of wasting opportunities buried in the data, making the wrong decisions about its handling or misinterpreting the insights is justified in the world which is as competitive as never before. However, it would be unwise to focus only on the technical problems of this explosion. Instead, it should be viewed as a challenge which brings opportunities of wide transformations in decision-making and knowledge creation.

The progress of computer technologies will probably solve the current technical problems associated with

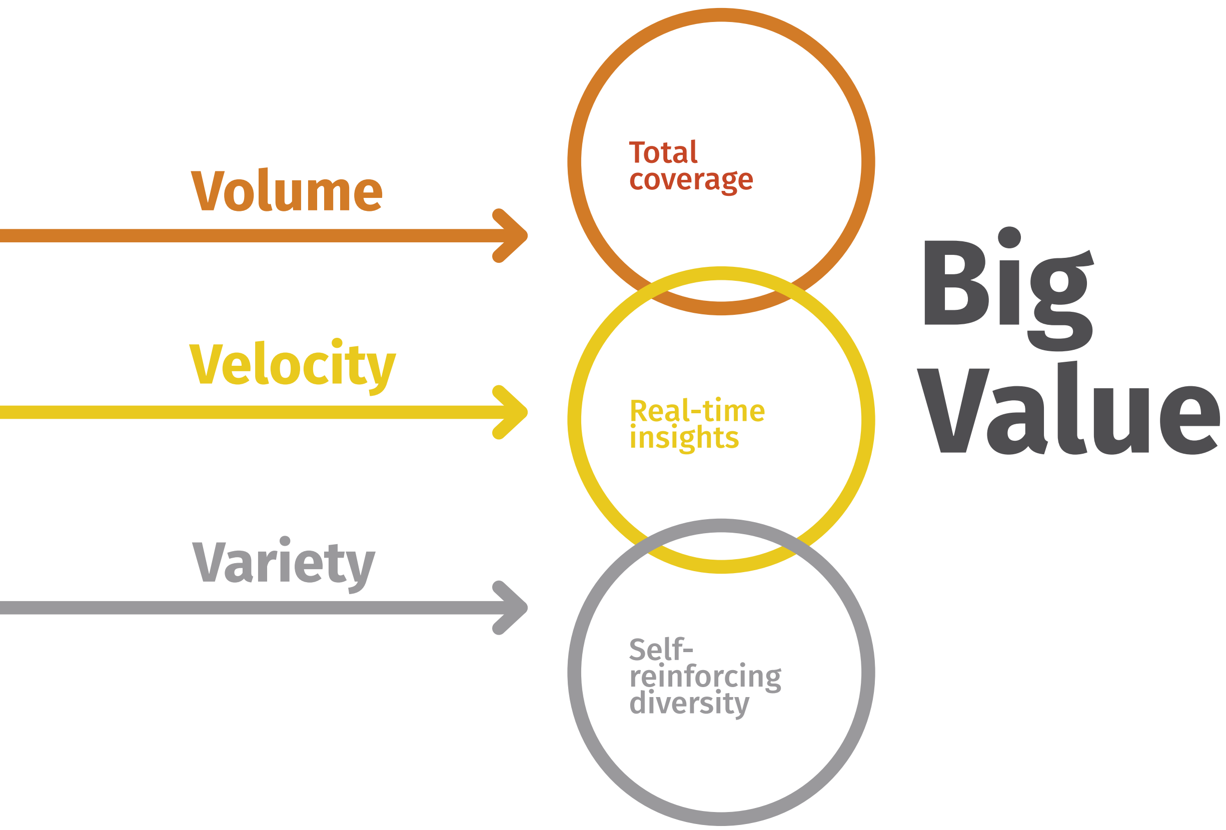
From **Big Trouble to Big Value.**

3 Vs transformation

Source: WISE Institute

data handling, however new challenges will appear together with even larger data flows. The frontier of what is *troublesomely* big is constantly moving, expanding the variety of the *valuably* big data. Luckily, while the troubles with handling data are of a temporal nature, the benefits and qualitative differences resulting from the growth of collected and analyzed information are permanent. Once you solve the technical challenges and shift from the disintegrated sporadic, small samples to real-time, whole data coverage integrated with other sources, the gains are here to stay.

The value of the data becoming really big is not simply a matter of scale. As data handling capabilities expand, the quantitative difference leads to a qualitative one. Coverage of the whole phenomena is not just a bigger sample. Real-time monitoring is not just an increased frequency of data collection. Ability to quickly and efficiently combine and re-use the expanding number of interconnected datasets is not just merging different sources of data. The fundamental difference and source of lasting value of big data is reaching the point where we get the *whole*, *live* and *expandable* picture of a given phenomenon. *Whole* data allows the capture of intricate aspects of the given process or study subgroups of interest. Live data shows detailed unfolding of the process in real time and gives us



a chance to immediately react to the unexpected circumstances. Complexity and numerous linkages between different phenomena mean that the data concerning one occurrence can provide important information about the other, so big data analysis is inherently *expendable*. The positive network effects occur: it is possible to gain fresh insights from constant building up, recombining and extending the scope of the analyzed data. Furthermore, all these improved insights can be used to make targeted predictions and decisions. The eye of the data scientist sees both a wider and a sharper picture than even before.

The tectonic shifts in thinking about the data handling, information collection and – fundamentally – perception of reality are already taking place. We have been used to living in a world of "small data samples", approximations, intuitions and guesses. Moving to the big data world allows us to see the whole picture increasingly accurately. It is like dropping the veil – improving both our perception of reality by means of increased data collection and storage and expanding analytical capabilities through using new data processing methods.

Another important change is the rebalancing between causation and correlation. In many cases it is sufficient to know exactly *what* is likely to happen given the particular circumstances (correlation) than *why* exactly it is happening (causation). Big data does exactly that – it gives us an increasingly exact description of what are observable outcomes of global events without explaining the underlying reasons. In principle big data allow for observation and prediction without a theory or model. The risk of following the inevitable false positives should be taken into account. The error-prone big data applications based solely on correlation may be suitable for many routine, repeatable problems like consumer profiling. However, the more complex analytical challenges (e.g. decision making support) should combine "old" approach based on theoretical models explaining the causation and tentative conclusions from small data analysis with the new opportunities for a deep dive into the big data pool provided by the innovative ICT solutions including Artificial Intelligence solutions.

Increased awareness of the real world phenomena helps us to make better-informed decisions and provides sound ground for research aimed at understanding the world. Still, it is only a tool. The decision on what to do with the gained insights remains a human responsibility, just as developing suitable theories to explain the phenomena observed through the new lenses. However big data should not lead to abandoning the responsibility and surrendering to the dictatorship of data. Instead, it should let people focus on decision-making and developing an understanding of the world without the need to struggle with approximations and false perceptions of what is really there. Data-based decision making and evidence-based policy are already widespread. The big data revolution will allow to ensure this is refined.

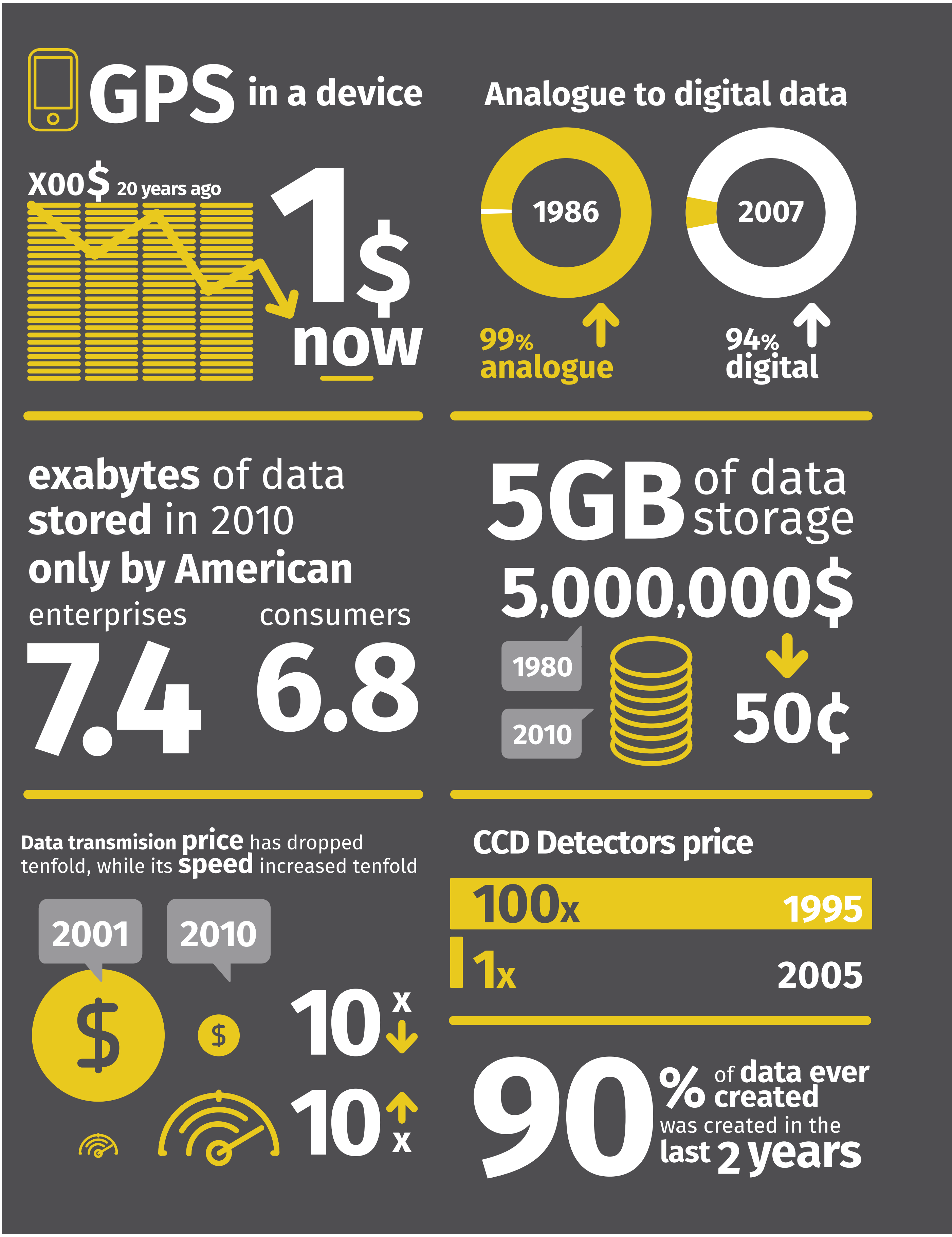
Think about an old Indian parable about blind men and an elephant. As each of them studied only one part of the animal they were unable to reach a conclusion abut the true form of the elephant. Transitioning from "small data" to "big data" is like letting these men see the whole elephant. They would still need an explanation why it looks as it does or decide whether it can be put to some good use, but at least now they would be busy solving the actual "elephant problem" instead of grappling with small portions of data about the animal.

Big data as another phase of the unfolding ICT revolution

The growing impact of big data on the economy is based on the breakthrough ICT tools which were developed during the last decade. However, it is the steady, exponential progress of a broad array of both hardware and software which made these breakthroughs possible in the first place.The most well-known evidence of exponential ICT improvements is Moore's Law. Formulated in 1965, it predicted a doubling of the number of transistors per integrated circuit every two years. What was initially a simple extrapolation of past trends over several years became a strikingly accurate prediction for almost half a century of ICT industry development. Initial doubling period was later reduced to 18 months as new technologies proved to be even more frugal than Moore initially thought, but the basic premise of exponential improvements holds true to this day. Thanks to technological advancements, the price of processors falls ever faster extending the frontiers of their viable uses in the economy.

However, the exponential improvements in the computer industry are not limited to processors. In fact, they characterize all of the key elements of the big data revolution – data collection, transmission, storage, processing. Take data collection, which benefitted from the falling prices of various sensors. Today including a GPS in a device costs around 1 $, while twenty years ago it was several hundred dollars. Falling prices and improving quality of CCD detectors used in cameras and scanners (100 fold price decrease between 1995-2005) allow for the digitisation of vast amounts of paper publications. The rise of the Internet and e-commerce – which in itself depended on advances in ICT technologies – provides an even more important source of new data for all aspects of our lives. These new data flows are huge and at the same time highly granular, which allows for the employment of insights gathered from the vast data pools to deliver highly customized services and detailed predictive analytics. The recent wave of datafication[4] resulting from the digital revolution overshadowed previous achievements in the field of data creation triggering speculation that it will be a similar transforma-

[4] Datafication means gathering and transforming information about the world into data which can be subsequently analyzed to gain insight about the reality. Although data was collected long before the ICT revolution, the switch from analog to digital allows to datafy the world at an unprecedented scale (see Cukier and Schönberger 2013)


GPS in a device
X00$ 20 years ago
1$ now
Analogue to digital data
1986
2007
99% analogue
94% digital
exabytes of data stored in 2010 only by American
enterprises
consumers
7.4
6.8
5GB of data storage
5,000,000$
1980
2010
50¢
Data transmision price has dropped tenfold, while its speed increased tenfold
2001
2010
$
$
10 x
10 x
CCD Detectors price
100x
1995
1x
2005
90% of data ever created was created in the last 2 years

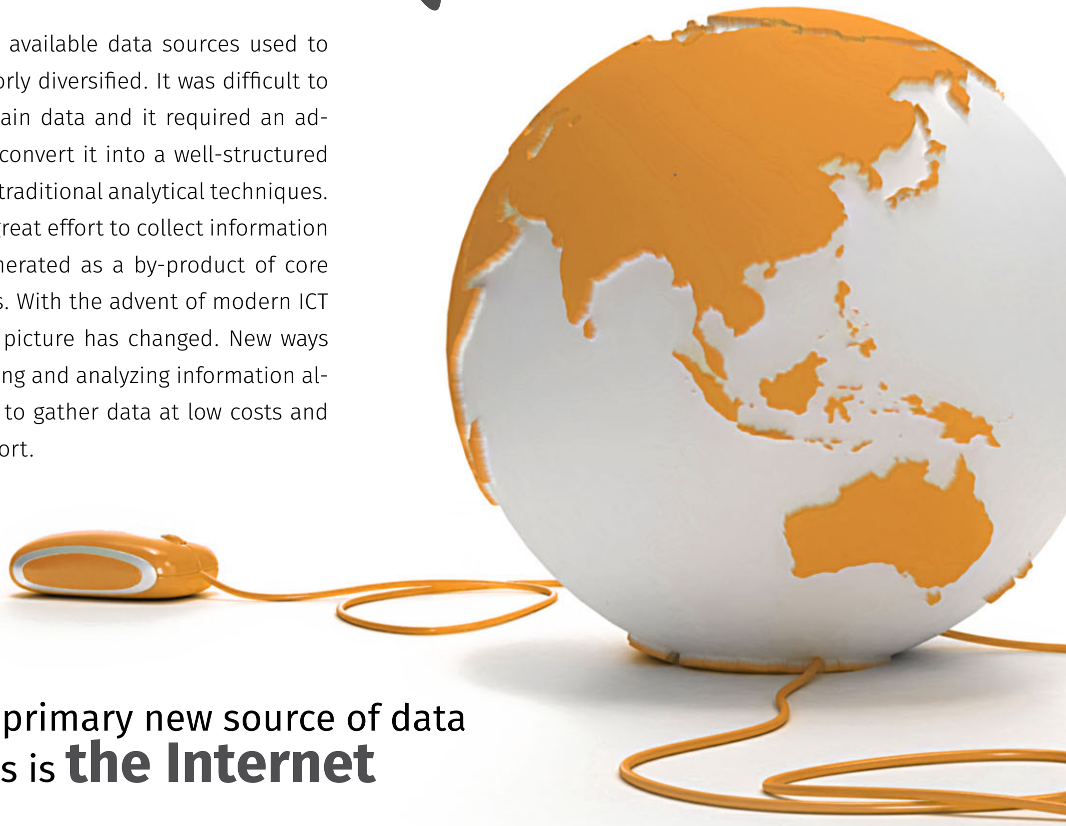
tional force for the economy, as was electrification a century ago. McKinsey Global Institute (2011) estimates that only 7.4 and 6.8 exabytes of new data were stored by American enterprises and consumers respectively in 2010. The world's technological information storage capacity has grown by 25% per year since the mid 1980s (Hilbert and López 2011) Moreover whereas in 1986, over 99% of all stored data was analogue, in 2007, 94% was digital. Today's world is undoubtedly digital. According to estimates by IBM (2012), the total amount of data generated today, amounts 2.5 quintillion bytes per day. As a consequence 90% of data ever created was created in the last 2 years. Even accounting for the rise of high-volume video and audio files, the data explosion is spectacular.However, these huge data inputs needed an appropriate network and storage infrastructure. Again, these fields have seen plummeting prices and maturing of important innovations in recent decades. The average price of data transmission, measured in megabits per second has dropped tenfold in the first decade of 21st century, while its speed went up by the same order of magnitude. The price of storage – arguably the primary cause of problems handling – has decreased, too. According to the trend observed for the first time by Mark Kryder (see Walter C. (2005)) storing 5 GB of data – an equivalent of a HD movie – would cost around 5 million dollars in 1980 and only 50 cents in 2010. As costs of storage plummeted, it became reasonable to keep every bit of data which businesses got hold of. Even if the possibility of given data having some value was low, its low storage costs in many cases justify storing, instead of scrapping, it.

Innovative algorithms are as important as better and cheaper hardware. By the estimates of Grace (2013), algorithms in six important domains improved the overall computer performance by 50%-100% over what had been previously achieved due to the hardware improvements. Hardware and software advances are complementary. Better equipment brings new possibilities and makes new software solutions viable, which in turn increase efficiency of its use. A good example is the rise of cloud computing and virtual machines. It was made possible by improved networks and availability of cheap, powerful computers which could be used for distributed computation governed by the new class of software. The introduction of these algorithms was viable because the necessary hardware was in place and at the same time new software allowed to tap into the full potential of physical equipment.

All these improvements are closely interlinked. For instance, improvements in processors and storage allowed for the creation of efficient pattern-recognition software which in turn led to data extraction from scanned images from books. Thanks to further advances in hardware, software and network technologies, this data then can be collected on an ever increasing scale, aggregated and analyzed at a relatively low cost and high speed, subsequently being sent to users around the globe. As the capability to collect, send, store and process data increases exponentially there comes a point when the quantity of analyzed data transforms into quality. Big data solutions emerge out of the rising tide of ICT progress. Thanks to exponential improvements, the further this process goes, the more profound the possibilities will be. Thus, big data shows the self-reinforcing power of ICT – the more it integrates into our everyday life, the more possibilities there are to use it productively. As big data emerges as a new, broadly applicable wave of innovations, it serves as yet another proof that ICT are General Purpose Technologies with long-lasting, transformational impacts on the economy.

Where is big data found?

Until recently the available data sources used to be scarce and poorly diversified. It was difficult to collect and maintain data and it required an additional effort to convert it into a well-structured form, suitable for traditional analytical techniques. It often required great effort to collect information even if it was generated as a by-product of core business activities. With the advent of modern ICT technologies this picture has changed. New ways of collecting, storing and analyzing information allow organizations to gather data at low costs and with minimum effort.



The primary new source of data flows is **the Internet**

It is possible to capture every action that a user takes and mine this so-called 'digital exhaust' for useful information, ranging from simple site statistics such as page views to sophisticated uncovering of online resources usage patterns. Detailed studies of the e-content use (websites, e-books, music, videos) allows companies to see what attracts users, e.g. how much time they spend in different sections of a web page, which parts of a video they skip, which chapters of a book they tend to re-read. Monitoring user activity is not the only way to gather data from the Internet. In the age of Web 2.0 and social media new big data flows have emerged – those provided by the users themselves. In addition to inferring consumer preferences from their interaction with websites, it is now possible to discern changing sentiments and new trends in real-time by studying user-generated content. E-commerce companies are an extreme case as their whole business is done on the Web where its every aspect can be captured and studied. Resulting adaptability and efficiency of this business model poses a threat for the traditional companies, forcing them to adapt to the new reality and embrace big data flows.

The spread of ICT in traditional sectors of the economy is another key enabler of big data revolution.

This is probably most profound in the finance industry. The digitization of money allowed for the recording of billions of transactions, analysis of financial markets in real-time and construction of increasingly detailed credit profiles for individuals. Another sector generating mountains of data is the health service. Data produced by monitoring patients and effects of different treatments meets the 3 Vs criteria. The health data flows are huge, they often come in real-time or at high frequency and they come in all forms, from text on prescriptions to images and videos from scanning. Finance and healthcare may be extreme examples of the data deluge stemming from the introduction of ICT, but similar processes are happening across the economy on a more modest scale. Digitization of office work and sales operations has created new data flows about employees' performance, effect of management practices, inventories and sales. ICT allowed brick-and-mortar businesses to hear their own heartbeats, even though they may be not as immersed in data as e-commerce. The bigger the company, the more data about its operations can be extracted, and the more valuable is the ability to aggregate and analyze it. In fact, this applies to all large organizations, first and foremost to governments. Public authorities have collected large amounts of data long before the age of digitization, e.g. during censuses, tax collection or welfare provision. ICT helps to handle these vast amounts of information and use them to find inefficiencies in public sector as well as to provide adequate evidence for policymakers.

The third main source of big data are sensors

Technological advances turned them ubiquitous and made the data easy to collect, especially through wireless networks. Continuous detailed measurements are possible not only during the production process, but also during product use. For example, an aeroplane producer may collect detailed data on the aircraft performance during flight. A similar practice may be employed for smaller devices connected to the Internet. Sensors in sold goods provide feedback which allows for improved and timely service as well as production process modifications immediately after problems arise. One particularly prominent new measurement technology is satellite navigation. GPS devices improve transport services and collect location data on numerous mobile devices allowing for the addition of spatial dimension to data analysis.

The three key sources of big data – the Internet, ICT in the workplace and sensors – are complimentary. As noted before, the value of data depends not only on its quantity and quality, but also on its interconnections with other datasets. Combining several data sources can lead to fresh, sounder insights, which increase the value of each particular source. One example of a technology overlap which creates a new valuable source of data is the emerging **Internet of Things.** The increasing use of goods which are characterized by a combination of an Internet connection and numerous sensors in everyday life means that by the end of the decade – when by some estimates (Evans 2011) there will be 50 bn devices connected to the Internet – a new dataflow about our homes and daily routines will emerge. These new devices will be accompanied by advances in infrastructure (e.g. smart meters in energy). Big data is entering our homes, bringing both opportunities, such as reduced waste, reduced energy use, time savings, but additionally, threats to privacy. Adoption of these solutions in the EU shall be fostered by the European digital policy agenda which is currently moving towards enabling a rapid uptake of the Internet of Things through standardization and research, development and demonstration while at the same time addressing governance and privacy protection issues (EC 2009, IERC 2012).

The fourth and so far unmentioned area which combines different sources of big data is **scientific research**

The so-called Big Science projects worth billions of euros in areas such as medicine, genetics, particle physics, astronomy or neurobiology generate mountains of data. Notable European examples of such projects include the ongoing CERN research (physics) as well as the recently launched Human Brain Project (neuroscience) or the planned Square Kilometre Array (astronomy) and European Spallation Source facility (material research). Scientists hope that exploiting this potential will benefit the public through improved productivity of research and by answering important research questions that could not be answered before. At the same time digitalization of research results and publication via Internet allow researchers to make use of a vast pool of relevant knowledge and observations. Combining insights from different scientific fields and their convergence is greatly facilitated by big data technologies, which on the other hand had to be stretched beyond state-of-the-art as the usual amount and complexity of the data related to Big Science outperforms the standards from commercial or public sectors. Therefore, both businesses and governments can exploit the experiences of Big Science to learn how to use increased data flows to their advantage.

How does big data affect the economy?

The business applications of big data are based on companies utilising new and improved insights. As information about a company's activities and its business environment are essential input for all businesses, the gains from better analytics have a vast number of applications across the economy and can be categorised in three ways:

resource efficiency improvements

Lack of information often leads to resource waste in production, distribution and marketing activities. Big data solutions (e.g. real-time sensor data analysis, advanced supply chain management, targeted advertisement) reduce the information gap and thus lead to efficiency improvements. Early adopters which are able to optimize resource inputs in value chain gain a substantial competitive edge. As innovations spread across the economy, the benefits for first-movers tend to decrease, while overall productivity rises, contributing to economic growth.

product and process improvements through innovation

Sources for innovative insights not only include R&D activities but also the day-to-day process of monitoring and consumer feedback.

management improvements through evidence-based, data-driven decision making

A better understanding of both the internal (strengths and weaknesses) and external (opportunities and threats) realities reduces uncertainty and provides a sound basis for both short- and long-term decisions concerning the enterprise.

The channels through which **big data affects** the economy

are closely linked to the abovementioned categories. There are three ways in which data can turn into valuable input for the company:

Tangible

Data-to-Product/Process

An effect occurs when insights from data analysis need to be implemented in the physical world to bring value to an enterprise. Implementing abstract insights into tangible processes and products may not always be an effortless process or generate instant benefits (e.g. data analysis may indicate the design weaknesses of a production line or the lacking features of a product, but there still remains the need to develop a fix in the product design or production process). However, this is still a rather straightforward way for industry and brick-and-mortar services to increase their efficiency (e.g. just cease deploying the resources where they are not needed) or introduce valuable innovations, especially as the data may point not only to the problems to solve, but also to potential solutions.

Intangible

Data-to-Information

This is the key source of big data value for business models where information is the core source of value. The main characteristic of this type of value creation is that all efficiency improvements and innovations take place entirely in the intangible part of the economy. The data is “mined” for valuable information and the value is created the moment the search succeeds (e.g. better investment strategy is found or useful insight on customer behaviour is generated).

Difficulty in gaining value

Data-to-Management

This is the most challenging way for companies to translate insights from big data to business value. It systematically brings data-based information into a company's decision-making process. A company not only needs to successfully mine the data and adapt the insights for management purposes, it also needs to adjust its corporate culture and move from instinct- to data-driven decision making. This source of big data value is open to companies throughout the economy, though data-intensive sectors with high rates of ICT adoption may gain the most.

These three channels affect different sectors of the economy disproportionately. Data-to-Information is especially important for the financial and insurance sector as its business models depend on gaining advantage from the information available on the financial markets and potential consumers respectively. The e-commerce and other areas of ICT-based economy are inherently dependent on maximizing the value of information available to them. Data-to-Product/Process will bring innovation and efficiency gains to manufacturing through better insights from various sources, including rapidly advancing sensors. Brick-and-mortar wholesale and retail trade can benefit from supply chain and inventory optimization based on big data analysis. Data-to-Management is broadly applicable across various sectors, but it can have the biggest impact on public administration and healthcare which are especially prone to inefficiencies due to lack of readily accessible information on performance and optimization opportunities. An additional impact for management in public administration and non-market services may be derived from open data, which will be discussed in the following chapter.

Overall, sectors potentially affected the most by big data represent more than a half of the EU-28 economy. Sectoral composition indicates that EU-15[5] may, comparing to the New Member States, expect a greater impact from Data-to-Information while at the same time a lesser impact from Data-to-Product/Process. However, much will depend on the rate of data-driven efficiency improvements in individual industries, public sector preparedness to embrace more data-driven decision making and availability of both human and ICT resources in a given country.

---

[5] EU-15 includes Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain, Sweden, United Kingdom.

The fastest uptake of big data solutions already is taking place among the sectors making use of Data-to-Information (primarily ICT and finance), as their business models are already data-dependent to a great extent. The next adopters will be companies drawing from Data-to-Product/Process which had always used some measures to improve their operations. Big data solutions will probably be adopted as another analytical tool offering better insights. The biggest challenge, and at the same time the biggest opportunity comes from Data-to-Management. Its potential impact is the broadest, but also the hardest to fully exploit. Sound judgment and a significant shift in thinking are required to take full advantage of the increasing number of insights from the available data. This effort results in significant performance improvements. As a widely-cited study by Brynjofsson et al. (2011) shows, companies adopting data-driven decision making are on average more productive by 5-6% than their intuition-driven counterparts.

The importance of a time component when pursuing big data projects varies between sectors and different parts of the value chain. The early adopter advantage is the key success factor in fields of data competition, particularly in consumer and social data collection through Internet services. Due to network effects leading to the domination of one or several biggest online platforms across all online activity, the successful Internet ventures become sole providers of their unique datasets with high value for other companies. Latecomers may find it hard to challenge the established players in the existing areas and may not be able to obtain a comparable amount of valuable data. Another area where the early adopter advantage plays a significant role is the provision of services based on big data. The ICT companies which will be first to develop novel big data solutions can gain access to client data, establish a strong presence on the market and benefit from learning-by-doing. Big data services latecomers will enter a market saturated with existing and well-known solutions. Replacing them will mean high costs and risks, creating significant barriers to entry for new players. This effect may be weakened to some extent by promoting the open data business models and big data solutions standardization, but it will still provide substantial benefits for early adopters. The importance of time will be less for companies outside the digital economy and the ICT sector as they either already have secure access to the data on their own operations or would need to pay for access to the data outside their core business activities. In this scenario, the early adopter advantage stems from the general positive effects of adopting efficiency-improving innovations ahead of competitors. These benefits will be weaker and more short-lived than in the case of the online personal data holders and the big data solutions providers, as their diffusion is much easier than in the previous cases. For example, there are no major barriers for the manufacturing company to adapt advanced sensors systems even if competitors were first to implement this solution.

>50% of the **EU-28 economy** is potentially most **affected by the big data**

5-6%

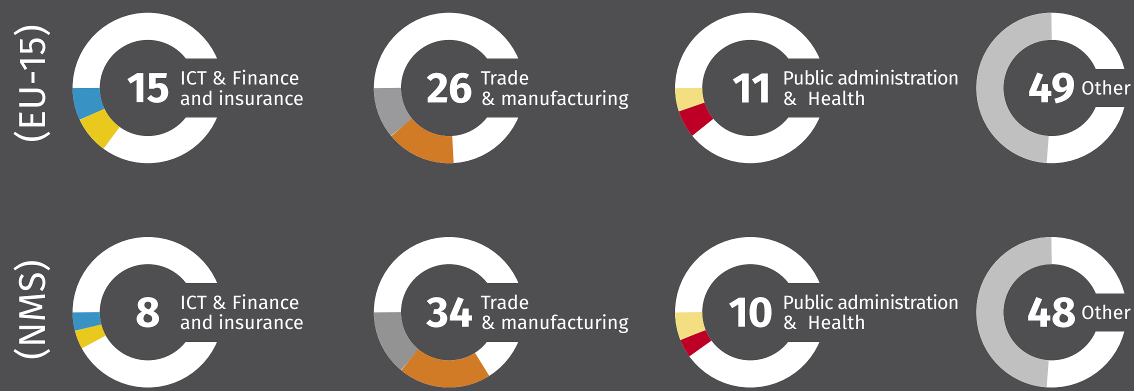
increased effectiveness with data-driven decisions making

Share in the EU28 GDP of sectors with **largest potential** influence of big data solutions **%**

Data-to-Information

Data-to-Product/Process

Data-to-Management



How might big data impact economic policies?

Besides the economic impacts of big data, one can speculate that its impact on economics and policy-making will also be significant, especially in the long run. The amount of data generated by economic activities is rapidly growing thanks to an increasing role of e-commerce and digitization of money flows. Some economists have already started exploring massive amounts of data created by the Internet economy to find new insights and verify established theories.

A Big data revolution offers opportunities for both microeconomic and macroeconomic studies. Online trade generates datasets which are both broad and detailed, allowing economists to come up with robust analysis of individuals' behaviour without the usual trade-off between the number of studied agents and the robustness of measurement. Additionally, companies selling their products online, especially on a large scale, are able to regularly conduct marketing experiments at low cost. This offers the possibility for cooperation with economists, who can conduct rigorous real-life experimentation, normally an extremely rare occurrence in the field of economics. There are already first examples of such projects, and the results are promising (see Einav and Levin (2013) for overview). It is also possible to "scrap" data from the Internet to track dynamics of macroeconomic variables. Notable examples of such an approach include the Billion Prices Project which provides real-time estimates of inflation based on data from online retail websites (Cavallo 2012) and research projects conducted under the UN Global Pulse initiative (UN Global Pulse 2012). The availability of real-time proxies and early warning indicators for key variables such as inflation, unemployment rate or output levels allows both governments and businesses to react faster to changing economic conditions. Instead of waiting for several weeks or months for new official statistics (outdated the moment they release), decision makers may use first signals of changes to prepare appropriate responses and use standard statistical data only to verify conclusions arising from real-time economic monitoring.

The collecting and analysis of big data for economic studies presented above is possibly only the first phase of an approaching datafication. Studying large data flows arising from e-commerce can provide a limited perspective on the economy, under representing offline activities of consumers and companies. However, as digitization of money flows increases, the picture of real economic activity arising from the analytics becomes ever more complete. Companies issuing credit cards already provide economic forecasts and "nowcasts" based on collected transaction data, while data generated by accounting software is also used to provide estimates for economic aggregates (Einav and Levin 2013). Financial ICT advances across the economy allow the monitoring of economic fluctuations on a daily basis at sectoral and subnational level, an impossible task with conventional statistical methods. These data sources are still not perfect representations of the economy, but they cover a much greater part of it than online stores. In fact, both the US and the EU already have to a large extent moved to economies, with cash used in only in 20 % of transactions in the US and around 10-25 % in the Northern European countries (MasterCard Advisors 2013).

One possible and relatively low-cost option for the near future is integrating all data about financial flows by modifying rules of data disclosure to key public institutions (e.g. central banks, ministries, statistical offices). Such economic data hubs – after appropriate anonymisation and aggregation – would provide policymakers with live detailed updates of economic activity and, more importantly, the ability to study and better understand the impact of economic policies. Many debates between economists arise from a lack of reliable data on the workings of the economy (consider current debates on austerity measures or decade-long discussions on price rigidities). Furthermore, as ways of doing business change, empirical findings from the past may not describe current economic processes well. That means that sound economic models of past may not work well today, given the changing nature of economic hypotheses. In this context new, more reliable and rich sources of economic data may provide invaluable insights regarding policy choices.

Even if all financial flow data was captured, the resulting picture of the economy would still be imperfect as there would only be information on gross flows and not value added, thus it would be difficult to estimate the actual role of particular companies in value chains. In the long run this too may change if moves towards universal digital accounting and currencies occur. As our ability to capture data on value-creation processes improve, an important data gap which hinders development of economics and evidence-based policy will gradually disappear.

However, even with such an “accurate picture” of the economy provided by data, researchers and policymakers will have to remain cautious. Aside from confidentiality and privacy concerns, which can be addressed by appropriate rules for handling the data and publishing the results of analysis, there are two important methodological caveats long-known in economic profession. The first is a Hayekian *notion of pretence of knowledge* (Hayek 1974). As authorities gain ever more data on economic processes, the propensity for extensive regulation and central planning increases . However, data may show only economic outcomes, without the underlying dispersed information which drives the actions of individual consumers and companies. No amount of raw data is enough without appropriate context and there are always limits to the knowledge of policymakers. Another important limit to data-based economic insights comes from *the classic critique of econometric modelling* by Robert Lucas (1973). Models evaluating policy options based only on correlations in historical data are prone to failure as individuals’ behaviour – and thus relationship between different variables – may change with introduction of new policy incentives.. Without finding the underlying, structural forces shaping individuals’ decisions, economists and policymakers will always be fighting the last downturn. The approaching era of data deluge means that careful modelling of true causes of shifts in economic activity and avoiding the pretence of knowledge will be critical.

Share of **cash transactions**

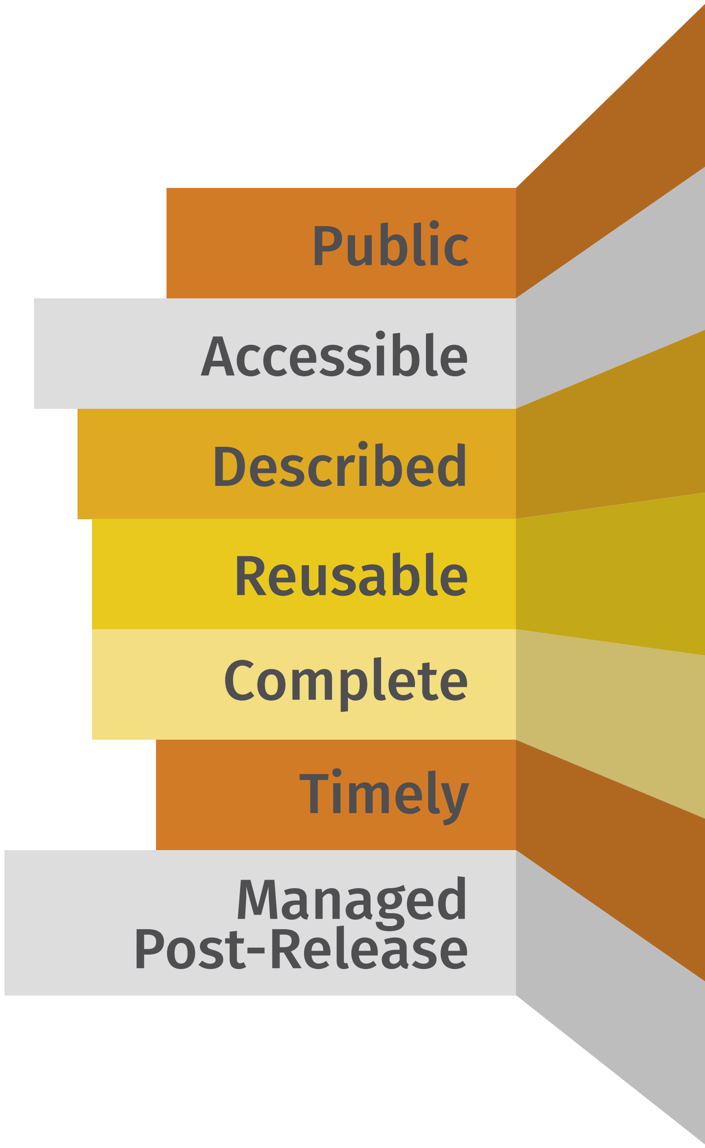
20% **USA**

10-25% **Northern Europe**

3 Open data

– old resources
but a new promise?

What is Open Data?



According to the Open Definition, a piece of data or content is open if anyone is **free to use, reuse & redistribute it**

— subject only, at most, to the requirement to attribute and/or share-alike. The main features of open data are feasibility and universality of free access and maximal freedom of use and re-use. This means providing data in ways that can be conveniently accessed by any interested party without additional barriers such as costs (price should include only reproduction costs, and in the case of dissemination by the Internet data access should be free), bureaucracy (no registration required), technological (data should be provided in open format) or usage restrictions (all forms of use and re-use, including commercial, should be allowed). Reaching the full potential of open data principles depends on how it is presented to the public. It is not enough to simply provide data for free. In the Memorandum on Open Data Policy released by the White House staff in May 2013, the key attributes of well-managed open data include (see visualization above):

shift of approach from "open by request" to "open by default"

data should be provided in formats that are as convenient, easy to analyze and modifiable as possible, with as few limits to use as the law permits

data is provided together with its description (metadata) to help users assess its strengths and weaknesses, as well as to know precisely what it describes

the data should be provided under open license which does not impose any restrictions on its use

the primary data should be published if possible and as close to its raw state as possible

to maintain the value of data – which often quickly drops with time – it should be released as frequently as possible, taking into account the needs of the key user groups

the contact point for the data users should be set, providing help and consultation, while at the same time receiving feedback which can be used to further improve access to the data

Another important attribute which is often overlooked in the open data discussion is the users' ability to verify **authenticity** of its source and its **integrity** (no external alterations to the originally released data), for example by the use of the electronic signatures. (Strancher et al 2013). This helps to decrease risks and costs associated with possible fraud and data mishandling.

Appropriate management of open data can be a challenging task, but it promises gains for both the data users and data providers – and not only in the public sector.

What is important about openness of data?

Opening up and ensuring easy access to data for all parties provides several benefits for data holders and data users alike.

Rediscovering the value & purpose of data. The nature of data can allow it to be used, reused, linked and recombined in countless ways without diminishing its value. Novel uses can be found long after the data was collected for its original purpose. Open data supports the emergence and realization of big data potential. While it enhances the volume and velocity of available data, its main impact is on the variety of sources. Opening data represents something more than reaching for the low-hanging fruit – it is more like finding new flavours in those already in the basket.

Fostering innovation. Providing maximum ease of access to data fosters experimenting and innovation. While data remains closed, it is impossible to assess its full potential or mine it for unexpected correlations and insights.

Mutual benefits – gaining feedback. The gains from innovative and more efficient uses of data are not limited to its users, but also benefit data owners, enabling them use the insights mined from the data to improve their operations. Opening data is a way to receive useful feedback, particularly important for public sector and research institutions. Citizens may receive more tailored offers as well as useful analysis of their data by specialized companies. Data is an asset, but the value comes from the insights which are extracted from it. Opening data may be a way to complement our own analytics by the insights from external sources.

Increasing transparency. Governments opening data yields an additional advantage of increased transparency, strengthened public control and enhanced public debate on their own actions. Providing open access to research institutions' data increases reliability of results by allowing easier replication of original analysis. Businesses may gain credibility and better relations with the public and authorities through providing data on their operations (e.g. showing sustainability efforts and their effects).

Network effects – the more, the better. Expanding the amount of open, linked, machine-readable data yields positive network effects as more elements can be recombined and experimented with. This is the snowball effect – each new open piece of data adds value to the whole ecosystem.

Where does open data come from and what are the incentives to open up?

The recent open data debate focuses on opening public sector data. However, governments are not the only possible sources of open data. Others include businesses, research institutions and individuals.

Governments collect a great variety and amount of data through operations of public entities. The main categories include data on activities of governments themselves, about businesses, citizens and the environment (in the broad sense, including research, geographic and resource data). Businesses collect data mainly on their own activities (business and research data) and clients. Research institutions collect data on the

Main **sources & types** of open data

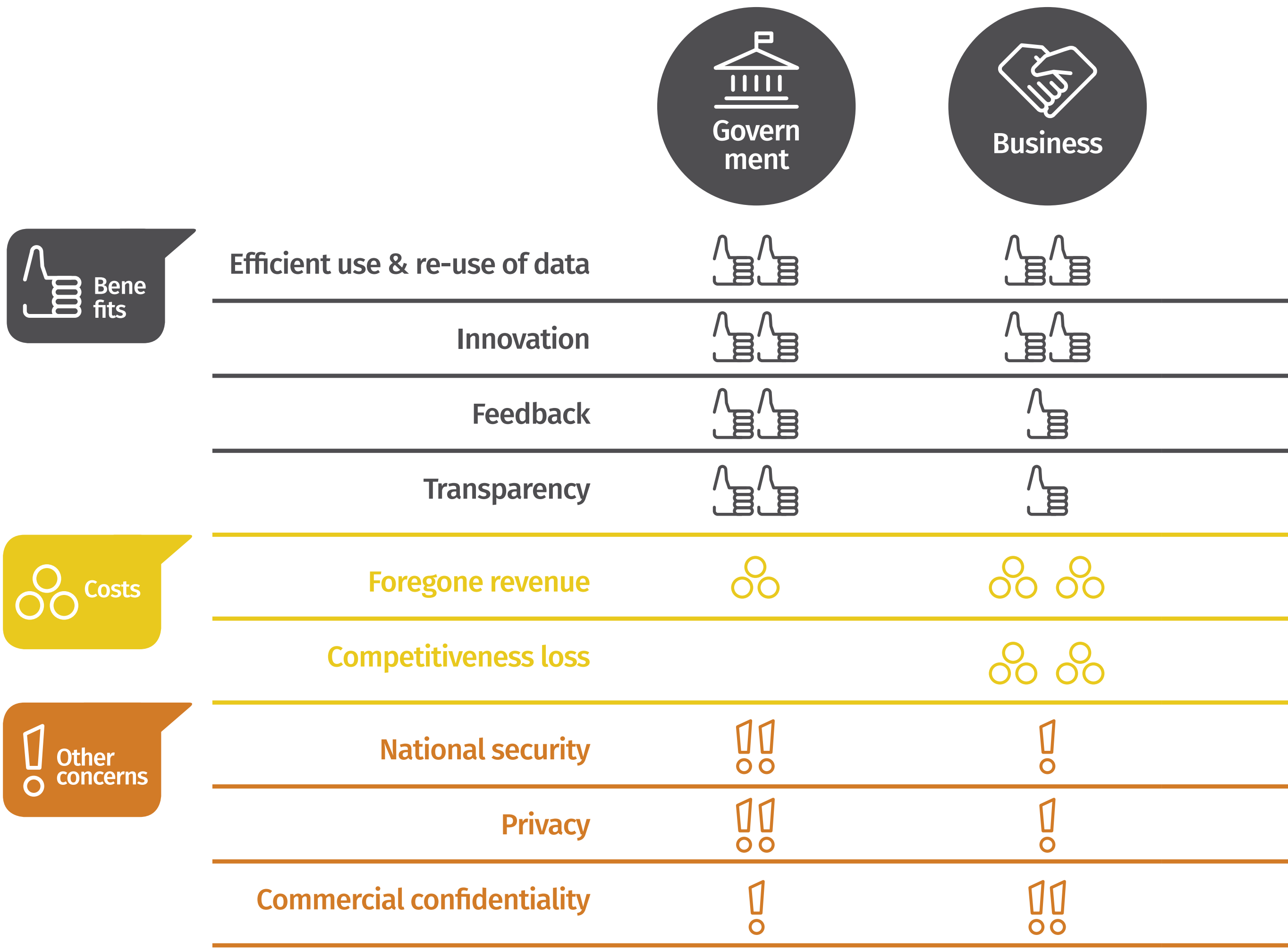
	Government	Business	Research institutions	Citizens
Public sector activity	++			
Private sector activity	+	++		
Individual activity	+	+	+	++
Environment description & research	+	+	++	+

Source: WISE Institute

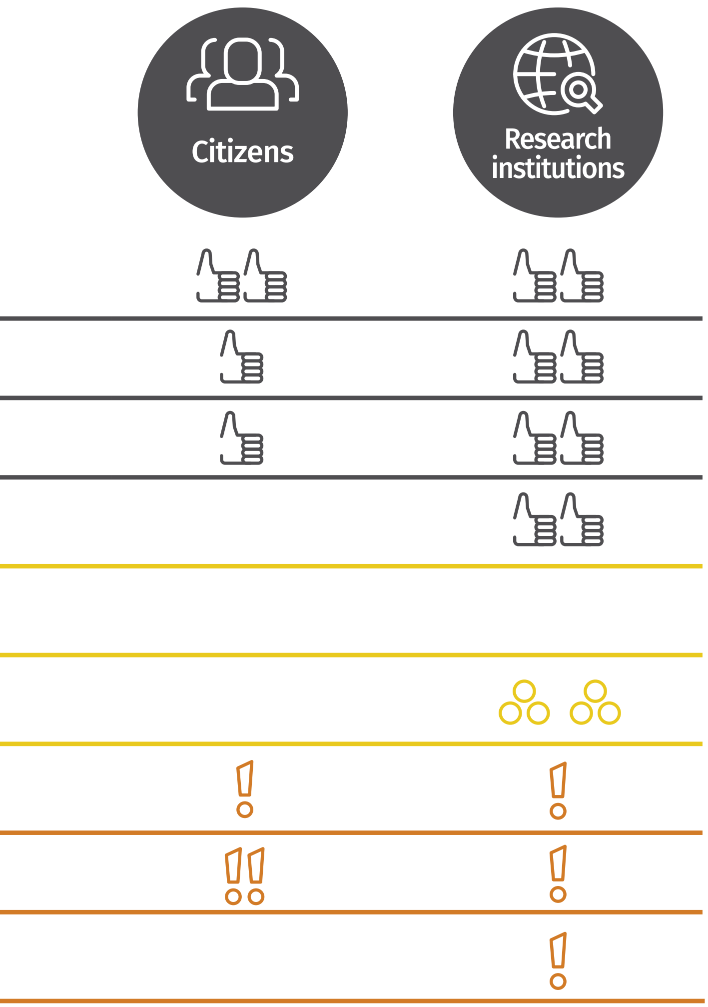
Costs & benefits

of opening data for its holders

Source: WISE Institute



environment and individuals through measurements conducted as part of research projects. Finally, citizens are typically able to provide data only about their own activities. However, this is gradually changing with the advent of portable electronics with advanced sensors and mobile Internet. Technological advances can now turn individual users into collectors of data not only pertaining to themselves, but also to their environment.



Falling costs of data storage and sharing facilitates are conducive to open data. The benefits of opening data outlined in the previous sections are now compared mainly with foregone revenue of selling access to the data and risks of losing competitive advantage.

Beside these direct cost-benefit considerations, there are three main reasons not to open data. Each of them corresponds to one of the three main stakeholders – citizens, business and government. The first issue is protecting citizens' privacy, which limits the use of personal data. The second one is maintaining businesses competitive advantages and ensuring fair competition by not violating commercial confidentiality. The third reason for not opening the data might relate to national security concerns. In such cases the data should not be released without the consent of the concerned parties. However, these concerns do not necessarily overlap with that of data holders. Governments collect data on businesses and citizens. Businesses collect data on clients and may hold data sensitive for national security (e.g. arms producers, cyber security providers). Research institutions may acquire personal data during research projects or their findings may be used in ways that threaten national security (e.g. biotechnology). Finally, even though individual citizens provide only small fractions of data about the total population, the aggregation and use of big data techniques may compromise national security. Thus, a legal framework protecting the interests of all concerned parties is required if maximum liberty and openness of data flows are to be achieved sustainably.

From positive to negative data freedom

The **key principle** of open data philosophy is **open by default**

This often means a paradigm shift for data owners. Instead of finding a reason why the given piece of data should be open for the public to use and re-use, they should start with the default notion that data should be open unless there are important reasons to restrict access to it. Thus, there is shift from what can be called *positive data freedom* (defining explicitly what can be opened) to *negative data freedom* (defining explicitly what should be closed). The latter encourages open data expansion and its innovative use by providing maximum access and freedom for data users.

Even if opening raw data may be not an option due to privacy, confidentiality or national security issues, there could be a possibility of providing partial access to data by aggregation and anonymisation (or pseudoanymisation). This is consistent with the negative data freedom approach – instead of closing data altogether, sensitive information is removed

from the datasets and the rest is opened for exploration by the public. However, the advent of big data makes it difficult or even impossible to successfully anonymise the datasets, especially those concerning individuals. As the numbers of datasets increase, it is becoming easier to combine them and de-anonymise the data by matching the overlapping records.

Another issue arising from the ICT progress is ease of data propagation. Once data is opened to the public, it is virtually impossible to reverse this action by ensuring that all copies are deleted. However, closing once opened data may be effectively achieved through banning its utilization by the users that are under the jurisdiction of the regulator. This shows the need for careful consideration of the long-term reasons for not opening data. It also indicates that shifting the emphasis from regulating data disclosure to regulating the data usage should be considered. Again, to ensure that the full potential of data is unlocked, data usage rules should be based on a negative rather than positive freedom paradigm.

The current debate about different aspects of data protection and dissemination seems to lack the holistic approach, a framework unifying and structuring the discussion about handling government, business, research and citizen data. The successful shift to a negative data freedom paradigm requires clear universal rules of when not to open data and which ways of using it should be prohibited. Without them, there will be only islands of openness in the ocean of closed data.

Share-alike licenses and attributing Open Data

The only requirements which are acceptable for open data are attribution (including clearly discerning obtained results from the original source) and the availability of using share-alike licenses. However, both options should be used with caution, as they can decrease the utility of the open data even without compromising its accessibility. In the case of attribution, it may be technically difficult to provide a full list of original data sources used if the data analysis included numerous complex datasets. The share-alike rule, which means that all derivative works should use the same license as the original data, is even more problematic. It implies that any work which used a given open data source should also be fully open to the public. This decreases the incentives to explore commercial uses of data as innovators would find it much harder to reap any possible benefits if they would need to be instantly provided for free. Not every business model can be built on the idea of open sharing. Share-alike requirements for open data effectively render some of its potentially beneficial uses unattractive. These caveats are common for all kinds of open content, but when it comes to data their implications might be especially problematic. One problem, already mentioned above, is the complexity of data analysis and its reliance on numerous sources which exacerbates any legal constraints created by additional open data requirements. The second one concerns the uniqueness of data. In many cases it is not possible to find an appropriate substitute for a given dataset as is possible, for example, with a piece of code. This means that the consequences of using share-alike licenses are more profound in the case of data, as they may leave the user no other way to achieve her goal than to accept the fact that her final work would be provided with an open license as well. The additional requirements described above can actually stimulate opening of the data, as data owners may use them as the tools to ensure that their data will not be used against them (e.g. a company provides research data to the public to get valuable feedback from innovators, but wants to discourage its use by competitors). However, if the objective of the open data provider is to maximize potential uses, it is better to provide data without any additional requirements, especially share-alike.

If the objective of the open data provider is to maximize potential uses, it is better to provide without any additional requirements, especially share-alike.

Open data from the public sector

Public sector data is at the center of the current shift to openness. This is not surprising, given a number of key characteristics which make it especially beneficial to opening up. Firstly, data usually comes as a by-product of other activities, thus costs of collecting it are already sunk. Secondly, the risk of losing the competitive edge in most cases is low or non-existent, as services provided by government agencies and other public sector institutions are typically non-market. As the public sector should be focused on creating societal value and increasing the prosperity of its citizens, any gains of third parties from opening its data should be viewed as consistent with its goals. Letting the private sector and NGOs make use of government data means leveraging their core competences of extracting economic and social value from the available resources. This allows the government to focus on its main goals without losing the potential of a by-product data it possesses. One of the implications of this approach may be summarized by the Tim Berners-Lee's call for "raw data now" – public sector should share as much data as possible without spending too much resource to refine it (as this could be done by third parties interested in identifying the hidden value). The key requirements to make data users' work better concern data format and dissemination mode, as summarized in the five star classification of data openness.

Opening up government data creates opportunities for economic benefits for both the public and the private sector as well as consumers. Open data can be used to create new business models or increase efficiency of existing ones. Public sector institutions can use insights gained by third parties to improve their own operations. For example, data may be used to identify neighbourhoods which require special attention due to social problems. Another example is highlighting key differences between lagging and leading schools or hospitals. In general, authorities can use output from open data users as an important source of feedback, a sort of mirror which can help them identify the areas where public services may be improved. Another important gain from open data applications is more efficient use of public sector services by better informed citizens. As open public sector data applications make information about public services more accessible in a user-friendly way, these services become better utilised and better tailored to societal needs. Applications comparing hospitals and schools or enabling public transport planning are examples of this effect.

Apart from the economic impact, there are significant non-tangible welfare gains from insights provided by opening public sector data. These include increased security stemming from better allocation of resources on crime prevention, better quality of public services and increased transparency for authorities which fosters public debate and strengthens civil society. However, it is important to remember about the limits of what open data can achieve by itself. The problematic issues can be technical (e.g. data collection biases) as well as institutional (openness does not translate into accountability if there is no effective feedback channels).

G8 Open Data Charter – Key principles
& best practices

During the G8 summit in June 2013, governments signed the Open Data Charter. It contains the commitment to open data principles and best practices which they are ready to promote.

1 Open Data by Default

- define your open data position in a public statement of intent
- publish a national action plan
- publish data on a national portal

2 Quality & Quantity

- use robust and consistent metadata
- publish and maintain an up-to-date mapping
- ensure data is fully described
- listen to feedback from data users

3 Usable by All

- make data available in convenient open formats

What about costs of opening data and keeping it open? Marginal costs of providing online access to data are close to zero, but the initial and fixed costs may be substantial. However, initial costs of collecting data were already covered to obtain it for the primary purpose or there were no costs as data was a by-product of some other activity. Costs of preparing data for publication are decreased by the fact that "raw" data is often acceptable to users who are ready to refine it to meet their needs.Finally, fixed costs associated with maintaining access to online databases keep falling with technological progress and the emergence of big data solutions.

The greatest cost of opening data is opportunity cost. As access to data may be sold, providing it for free means that potential revenue is lost. However, while foregone revenue is a cost for the public sector, it is beneficial from a societal perspective. If extracting value from given data could only be done by big players ready to pay the price, then opening data would simply mean a transfer to these companies. However, as the information industry is a highly com-

5 Releasing Data for Innovation

- support the release of data using open licences
- ensure data is machine readable in bulk
- release data using application programming interfaces (APIs)
- encourage innovative uses of our data (challenges, prizes or mentoring for data users)

4 Releasing Data for Improved Governance

- develop links with civil society organisations and individuals
- be open about our own data standards
- document our own experiences of working with open data

These principles and best practices can be applied not only to G8 governments, but also serve as a useful benchmark for other countries and regional authorities. They are also applicable – after some changes – to businesses and research institutions. For example, it is equally important to publicly formulate the open data position, have an action plan and one place to publish data for the private corporation as it is for the government. In general, all types of big organizations – public, private, research – may learn the best ways of capturing the value from opening data from each other.

petitive sector, small developers also can make good use of data and giving them access to it reduces barriers to entry additionally imposing competition on the market. Furthermore, thanks to the possible value gains from combining different data sources, opening data brings broad, marginal improvements across the economy which would be lost if the price for secondary uses of data was the same as primary use. The additional gain for society are innovation opportunities stemming from the large pool of data provided for free to anyone willing to explore it for new valuable applications.

When seeking to maximize value from opening data, it is important to focus on the open data "marketplace" as a whole – that is, to take into account both the supply and demand side. While ensuring free access to data provides the supply side, there may be a need to foster demand by increasing awareness and skills of potential data users. Public sector may take a step towards the data user community and provide a framework for exploring possibilities hidden in the data. It can help to organize events and contests to bring together creative individuals and companies.

Open data – **global good practices**

Increasing use of public services

While spatial data from satellites has been open for a long time and is widely used for improving transport efficiency, open data on public transit is a relatively new phenomenon. The open transit data is already available for many cities in the US and Europe. By providing data free of charge, public authorities spur competition between app developers which in turn brings more useful solutions to the market. The transit apps allow city dwellers to easily plan their trips on public transport, increasing the use of this service and reducing traffic. For example, the wide range of transit apps for the US may be found on the City-Go-Round website which helps users choose the app most suitable for their needs.

Improving transparency and quality of public services

While a lot of data on the performance of different public services providers is collected by public authorities, it is often dispersed and not easy to interpret for citizens. Opening up this data and letting third parties present it in an attractive and coherent way, solves this problem. The Guardian GCSE schools guide provides an example of this approach. It is based on open data about the UK schools'performance in GCSE exams. Parents can easily find relevant information about schools in their neighbourhood and compare available alternatives. The project might also increase public awareness of the challenges faced by the education system.

Commercial application – fostering energy efficiency

Husets Web is a Danish website which provides citizens with information on energy saving measures appropriate for their homes. Recommendations are based on public data about the house, nearby energy projects, financing possibilities in a given municipality and possible contractors. The service is free to use. The company's clients are municipalities which are interested in reducing energy consumption and stimulating local economy.

Sources: citygoround.org, theguardian.com, ourservices.eu, Piketty and Saez (2003), healthdatapalooza.org, odi.org.uk, midatalab.org.uk

Research based on public sector data

The study of income inequality in the US based on IRS data is an example of the importance of public sector data both for science and public debate. The research began with a seminal paper by Piketty and Saez (2003) that sparked economists' interest in the divergence between productivity and wages growth in the US.

Demand push

Demand for open data might be stimulated by special events dedicated to hidden value search within data from a specific area. U.S. Health Datapalooza is an example of such an event. It is an annual conference which focuses on innovative tools based on open health data. Stemming from governmental efforts to open health data, it stimulates knowledge sharing by collecting experts from numerous fields connected to health open data and provides incentives for innovation such as competitions and hackathons focused on finding new ways of utilizing available data. Similar events on health and other issues may be organized in other countries as a way to magnify the impact of opening data in areas of special public interest.

Raising awareness of open data opportunities

Another way of stimulating the demand for open data is supporting institutions that raise awareness of open data opportunities. For example, the UK government supported the Open Data Institute which stimulates the uptake of open data solutions in the private and the public sector by bringing together experts, collecting and providing knowledge and advocating opening up data.

Public-private partnership for promoting responsible use of consumer open data

The business accelerator midata Innovation Lab is run by the UK Department for Business, Innovation & Skills with support from private enterprises. It focuses on offering private sector the opportunity to develop consumer-focused applications based on personal data. The data comes from volunteers' personal data stores. The lab cooperates with the consumer protection bodies to find new solutions which empower and protect consumers while creating new business opportunities. For example, a customer may upload data about her mobile phone usage to her personal data store and then allow an application provider to use it to determine the contract which best suits her needs.

Public sector open data impact – European perspective

Despite the common rules for providing the government data set by the Directive on the re-use of public sector information (the PSI Directive), EU countries show very diverse commitment to opening data. The scoreboard published by the European Public Sector Information Platform provides a synthetic measure of the status of open data in the EU. The UK is leading the transition to open government not only by opening up ever more public sector data but also through stimulating the uptake of the solutions based on provided data (e.g. midata Innovation Lab, support for ODI – see the box on global good practices).

Beside the UK, European open data leaders come from both the North and the South of the EU-15, showing a lack of North-South divide in this respect. However, the New Member States are substantially lagging behind in the uptake of public open data. The cross-country differences are explained mainly by the additional stimulation of the demand for open demand and access to regional data. Together with UK's example, this indicates the importance of expanding public sector involvement in an open data movement through support of open data application uptake and soft measures such as fostering knowledge transfer. If European countries want to reap full benefits of timely opening up, simply opening data may not be enough. Creating a friendly environment for the emergence of open data solutions is needed. This observation is important in the context of the current revision of the PSI Directive which secured the key principle of negative data freedom across the EU. Though an important step, it should be followed by active participation by the public authorities on the European, country and local levels.

To measure the potential impact on the EU economy of opening public sector data, a ranking of sectors most conducive to open data was developed on a Deloitte (2013) study on data.gov.uk experiences. As there is yet no information on the intensity of actual usage of particular datasets by each industry or the inherent added value, the ranking should be used for illustrative purposes rather than for quantitative assessment. Nevertheless, by grouping the sectors into several categories and comparing them to the economic structure of European countries, several tentative conclusions can be made.

Open data from the public sector can be broadly applied in around 10% of the economy, and substantial effects may materialize in another one third of it. Overall, there aren't many businesses for which improved access to public sector data would not offer meaningful benefits. Another observation is the link between economic development and the potential impact of open data. As knowledge-intensive and social services amount to a higher share of GDP in the EU-15 than in the NMS, the immediate impact of opening public sector data will be higher there. On the other hand, open data may provide an additional impulse for and speed up growth of value provided by services in NMS, thus fostering convergence within the EU. Finally, unlocking value of the public sector may help sectors troubled by recent financial and public debt crises (finance, real estate, public administration) as well as better prepare Europe for long-term challenges of an aging population (health services) and building a knowledge economy (knowledge-intensive sectors).

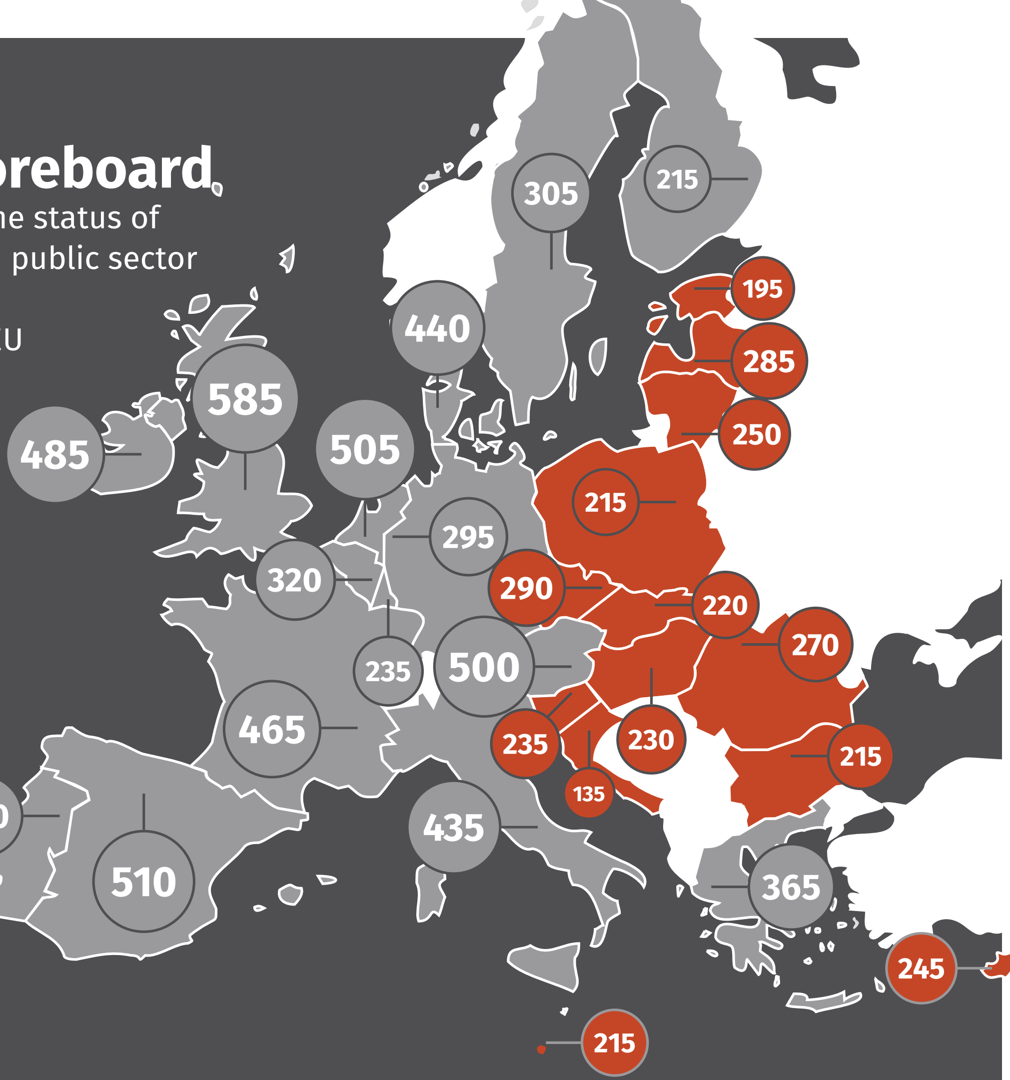
ePSI scoreboard

– measuring the status of open data and public sector information re-use in the EU

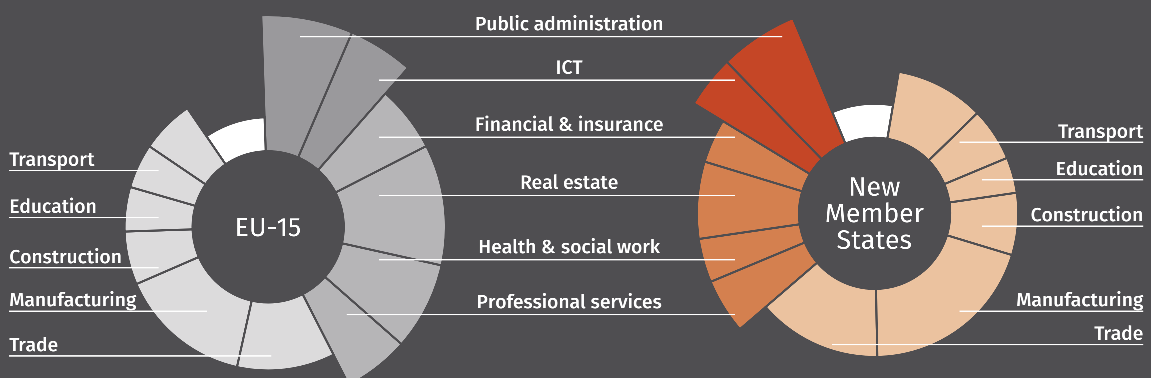
Source: European Public Sector Information Platform

EU-28

321



Share in the EU28 GDP of sectors with **largest potential** gains from open data



Source: WISE Institute based on Deloiite (2013) and Eurostat data

4 **Modelling the economic potential**

of big and open data
in Europe up to 2020

Despite a recent surge of publications on the potential of big and open data to spur economic growth, there are only a few studies which attempt to systematically quantify this impact at a macroeconomic level.

Two notable examples are McKinsey Global Institute (2011) report on big data perspectives in several key sectors (manufacturing, retail trade, transport, healthcare, government agencies) and Cebr (2012) study on potential impact of big data on the UK economy based on in-depth literature review and economic modelling of secondary effects. Both texts focus on a bottom-up analysis. In other words they start with sectoral impact assessments and try to aggregate them to the entire economy. An alternative approach – top-down estimates of key economic variables, based on macroeconomic type of reasoning – may provide only illustrative results that are usually based on generalized assumptions about the future. Since the applications of big and open data in real businesses had started to emerge relatively recently their scale is rather limited. As a consequence there is no macroeconomic evidence if their future impact on the economy will be large or small. Thus, top-down attempts to quantify the effects of new ICT trends are usually based on ad-hoc assumptions taking into account the relatively few observables like the documented productivity surge observed in the 1990s[6] Nevertheless, such estimates may serve as a useful benchmark for assessing the results of bottom-up studies, putting them in a broader context. The same could be said about the historical examples provided by the macroeconomic results of past proliferation of such general purpose technologies like steam engine or electricity.

In case of electrification productivity started to increase faster because the old – steam powered – machinery was replaced by more productive electric engines (David 1991). Moreover, what was important is not the general access to more power but rather a larger flexibility and adaptability of electric motors that resulted in increased productivity growth. This analogy can be applied to the big and open data economic potential. We can look at it as an aspect of a broader ICT revolution. Although, the initial macroeconomic impact of ICT proliferation was probably negative (see Jovanovic and Rousseau 2003), in 1990-2008 we observed large acceleration not only in GDP but also in labour productivity growth, that could be attributed to the proliferation of ICT technologies to most of the sectors in the economy. In the current decade the second wave of this revolution can be expected if the low-hanging fruits of big and open data are picked (e.g. improving efficiency through noticing and cutting the resource waste, gaining value from relatively easy to construct, but detailed and wide indicators of customers' preferences), and once again not the direct productivity gains will be the most beneficial economically but rather the indirect impact that can result from flexible applications in various industries. given on the analogies drawn from the electrification era (1900-1940) one can expect an additional increase in productivity through broad big and open data usage by 1-4% (see Edquist and Henrekson 2006). This number should however be treated with caution as some (Jovanovic and Rousseau 2003) perceive the ICT revolution as more dramatic and promising than electrification, whereas others look at it as a rather temporary and short term phenomenon.

---

[6] For an example of this approach, see the report on the Industrial Internet by Evans and Annunziata (2011)

focus on 3 key forces

To arrive at robust estimates of big and open data impact on the EU, we focus on three key forces which should affect the economy on the aggregate level and incorporate them into the bottom-up model of the EU economy. These effects are:

productivity increase in manufacturing & services

resulting from increased business processes efficiency, better supply chain management in retail trade and fraud/error detection in finance, insurance and government agencies. These are associated with big data advances.

competition improvements

resulting from lower entry barriers for business as a consequence of the opening of public sector data,

improved allocation of production factors

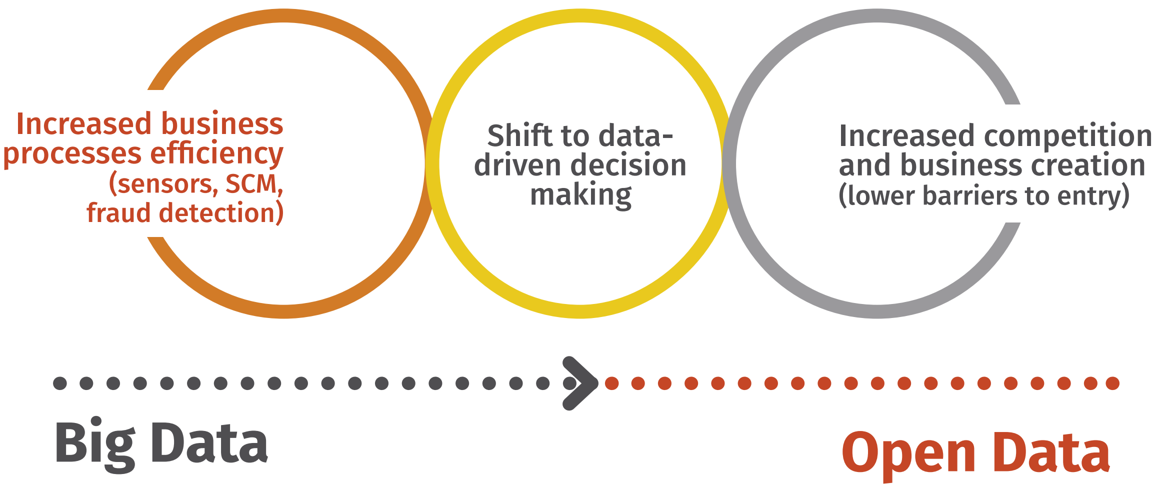
resulting from better decision making due to a shift from instinct – to data-driven management processes. In most sectors better allocation should be associated with big data, but in some cases (government agencies, healthcare, education) open data is also expected to play an important role by providing additional incentives for management improvements.

Our bottom-up, projection model (which we call BOUDICA or **B**ig and **O**pen **D**ata **U**niversal **I**mpa**c**t **A**ssessment model, see: Technical appendix) reflects the changes to value added in 21 sectors for every European country, each with a different pace of adoption of big and open data innovations. This difference is modelled on the basis of data on the current ICT adoption by European businesses in different countries as well as literature review of different sectors' propensities to adapt big data solutions or to make use of open data. Additionally, BOUDICA takes into account the size of enterprises in various sectors and countries as this affects their ability to pursue big data projects. The BOUDICA model incorporates detailed data on European economy extracted from the Eurostat database. A detailed description of the BOUDICA model can be found in the methodological appendix.

Several important conclusions arise from the BOUDICA model projections. Firstly, the primary po-

Key macroeconomic effects

of big and open data introduction



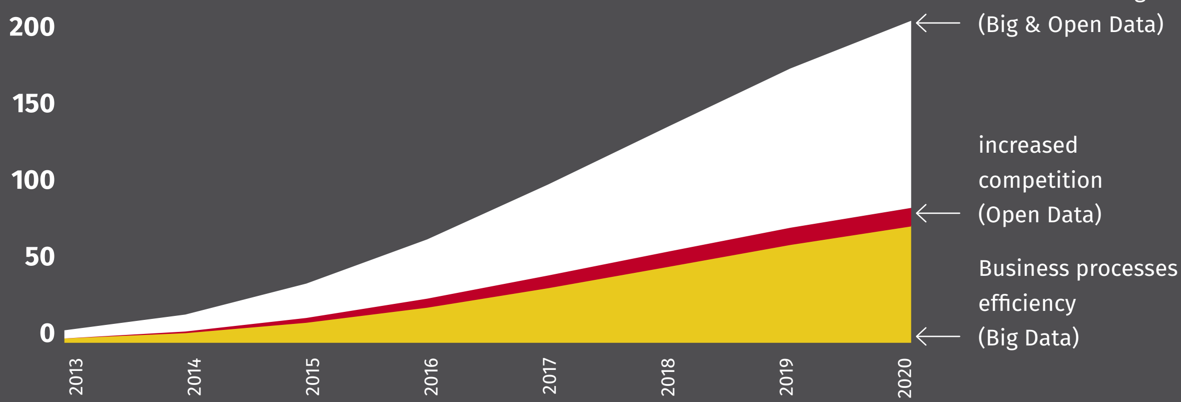
tential of big and open data still has not been exploited to date. In fact, as businesses only begin to grasp the idea of economic utilization of these new sources of value, current influence of the data deluge on the developed economies is only a small fraction of the likely long-term gains. According to our estimates, by the end of the decade the total annual benefits for the EU economy from big and open data may surpass 200 bn EUR. Secondly, the greatest potential in terms of productivity growth lies in Data-to-Management but its realization remains conditional on the shift to data-driven decision making in European businesses. Exploiting economic opportunities in this area to the full will demand a substantial organizational effort as the transformation of the corporate culture usually lags behind the technological progress. Expanding access to open data should foster this shift, especially in the public services sector. Productivity of European companies may also rise substantially due to the improved business process efficiency. Reduction of operating costs through better identification of production malfunctions in manufacturing and fraud or error detection in the financial sector and public administration should play a key role in this case. Big data potential in this area is probably much easier to exploit, especially for large corporations that already have enough skills and capital to successfully grasp the opportunities arising from big data applications within the next decade. With respect to small and medium enterprises that are more financially and operationally constrained, quick deployment could be more demanding and therefore significant savings in costs will probably be achieved later in time. We estimate that a direct effect of an increased competition stimulated by the opening of a public sector data will be much smaller than the broad productivity increases based on big data. Still, annual benefits will be substantial, surpassing 10 bn EUR by 2020.

Estimated **benefits** by 2020

200 bn € for the EU economy **from big & open data**

10 bn € from opening **public sector data**

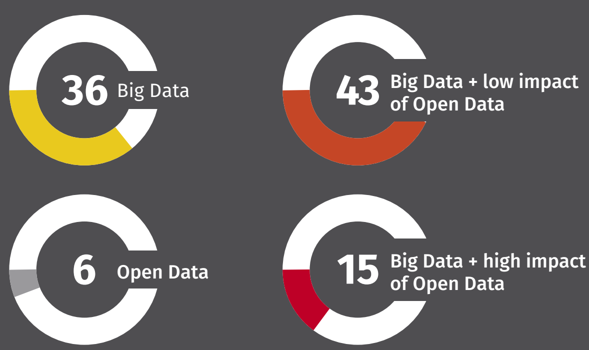
Additional **annual** GDP in bn € in the **EU-28**



Additional GDP **by source of impact**

Economic potential of big and open data

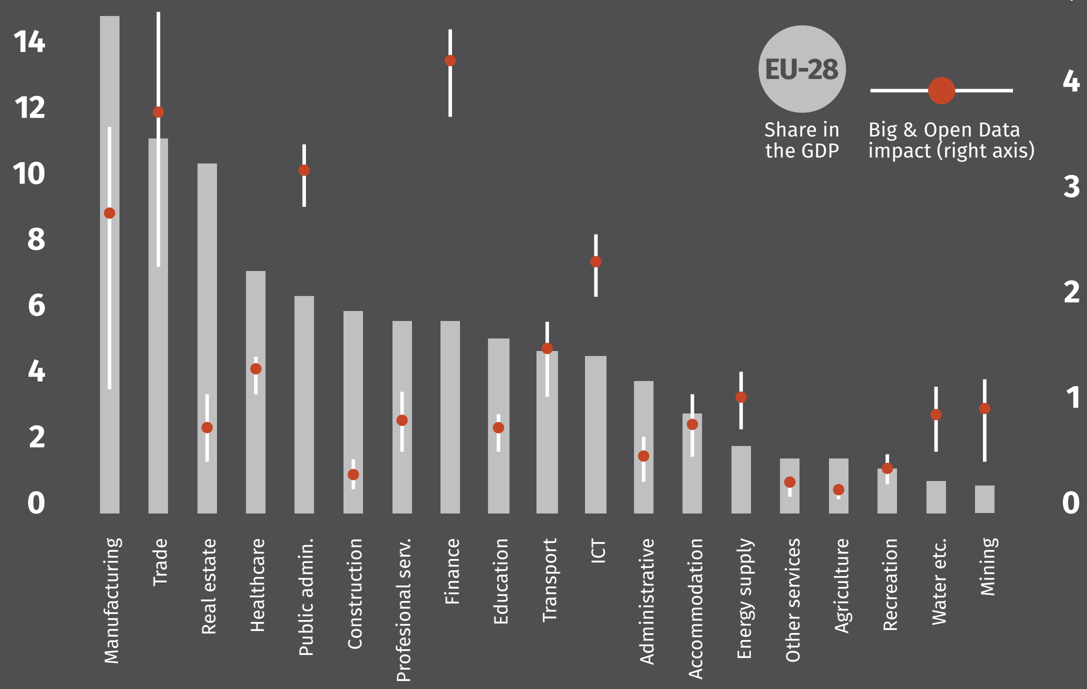
%



Economic potential of big and open data – Additional GDP **by sector** in which data-driven solutions are introduced %

Economic potential of big and open data – **impact by sector**

% (the EU average impact & its range for individual countries)



Overall, by 2020

big & open data can **improve the European GDP by 1.9%**

an equivalent of one full year of economic growth in the EU

The majority of this effect will depend on big data applications in management and resource efficiency. The direct macroeconomic effect of open data will be relatively modest, but it will play an important role as complimentary element of the big data revolution.

Sectoral Impact

The impact of big and open data will vary greatly according to sector. Applications of data-driven solutions in manufacturing, trade and logistics should account for almost half of total big and open data impact, whereas ICT, finance and public administration will deliver another one third of it. Differences in the sectors' contributions to the overall growth come from their different,relative size, distinct economic features and technological characteristics. The biggest data-driven transformation can be expected in the financial sector, trade (including e-commerce), public administration, manufacturing and ICT. On the other hand, sectors traditionally dominated by public financing like healthcare, education or municipal services can expect a much smaller, although still not negligible returns from investments in big and open data projects. This observation contrasts with the public administration that could potentially belong to the largest beneficiaries of the big and open data revolution. Among the largest sectors in the European economy only construction is not likely to see substantial gains from big and open data. Outcomes vary not only between individual sectors but also between countries, depending on company size, structure and general level of ICT penetration. Countries which are currently lagging behind in ICT solutions uptake and whose industries are dominated by small and medium enterprises will find it more difficult to extract economic value from big and open data. In such cases lower adoption rates of data-driven innovative solutions and management practices will result in an overall smaller positive economic impact.

The relatively modest contribution of the ICT sector in the simulations comes from the fact that the available statistics show the macroeconomic effects of big and open data applications in particular sectors and not necessarily the growth of value added within the sectors induced by the innovation. In fact, the ICT sector will see substantial gains from investments in innovative data solutions in other sectors and will capture part of the increased value generated by data-driven productivity increases. For example, applying big data analytics to supply chain management in a retail company will increase its productivity, but at the cost of necessary investments in the ICT infrastructure and software. At a macroeconomic level this cost turns into additional value delivered by the ICT sector and overall impact on the economy reflect total productivity gains provided by big and open data. The size of this effect depends on the degree of monopolization of the big data solutions on the providers' side.

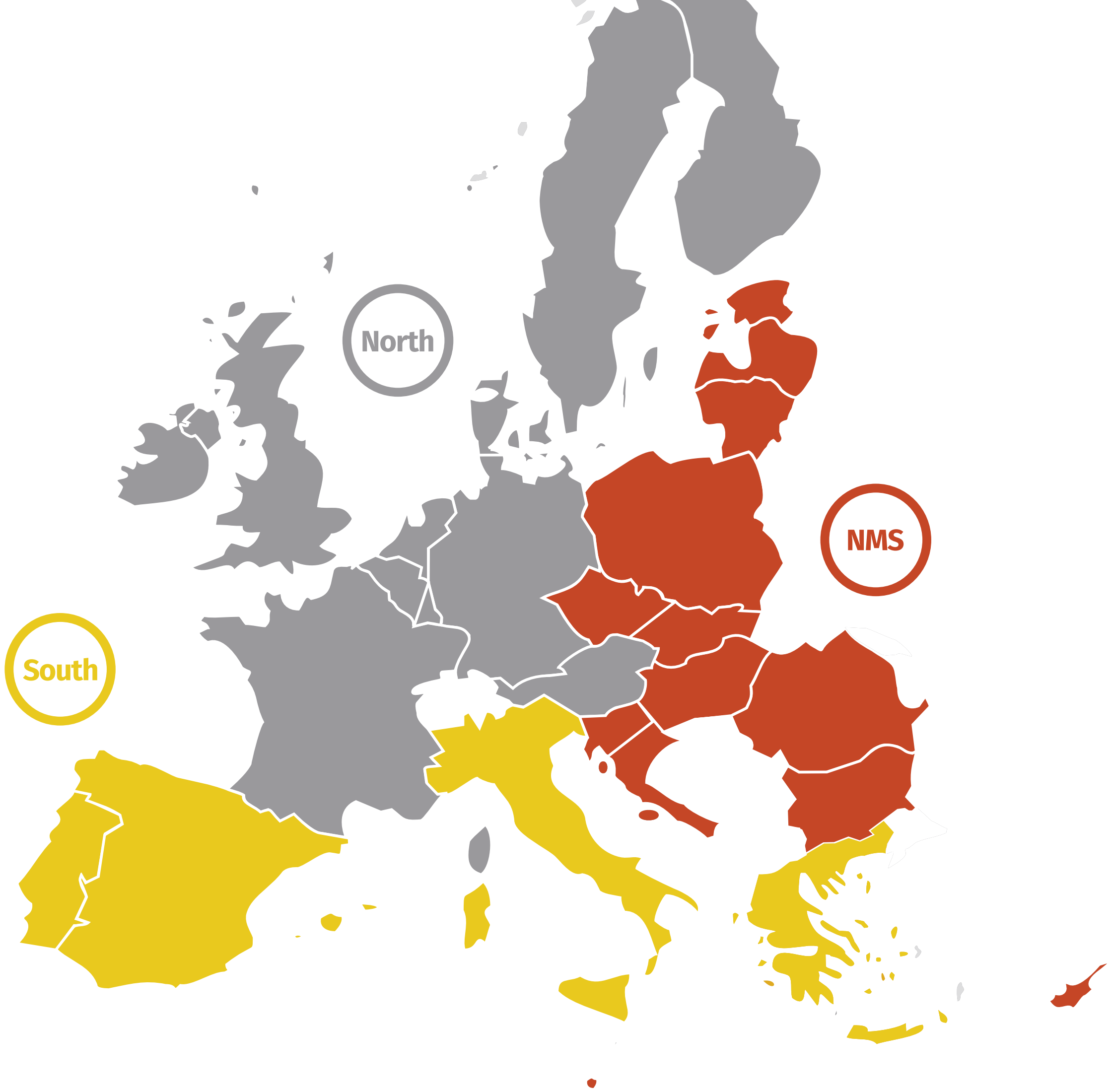
Another important implication of the value transfer between the sectors is the dependence of the final macroeconomic outcome on the competitiveness between local big and open data solution providers (data owners, analytic hubs, parts of the global ICT supply chain). For net exporters of innovative ICT solutions, the positive impact of big and open data on GDP will be greater, while for the net importers it will be less. This provides an important rationale for the EU to invest in the development of the competitive data analytics industry. It should be noted that – just as in the previous phases of the ICT revolution – the main impact on the GDP will stem from the increased productivity in the non-ICT sectors rather than additional value captured by the ICT companies, so that imported big and open data solutions will still be more beneficial than no solutions at all. For the purpose of the presented modelling exercises, the economic potential of big and open data by country does not include net exports of the ICT sector related to this area.

Structural differences between countries play an important role for the diversified projected macroeconomic impact of big & open data revolution amongst EU member states.

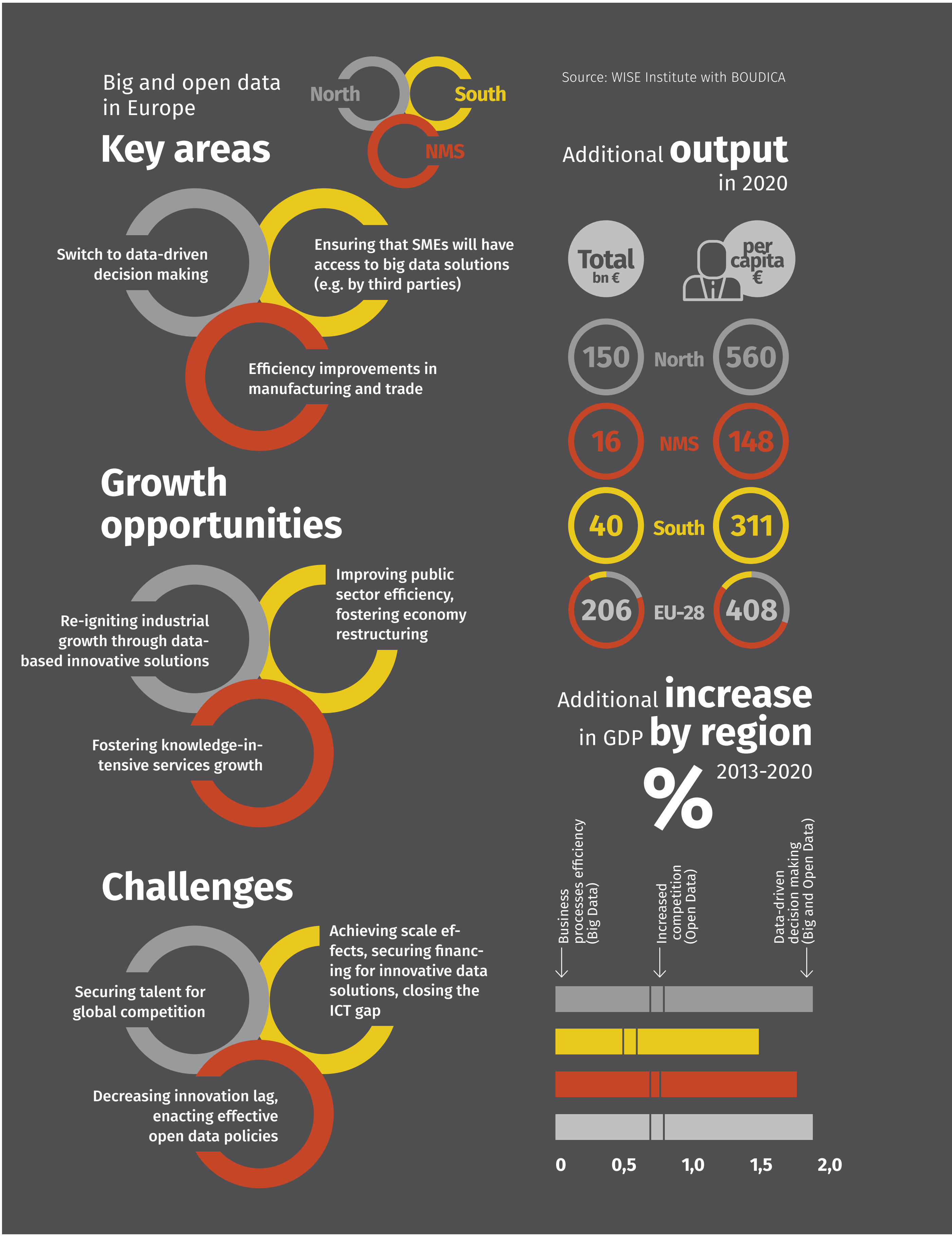
Country-specific differences

The BOUDICA projections suggest that in **Northern Europe** – except France – we can expect a stronger than average increase in GDP level due to exploitation of the economic potential of the data. On the other hand, most of the **New Member States & Southern European** economies should benefit significantly less from this technological revolution than the rest of the continent. Significant exceptions are the **Czech Republic & Poland** – the first country may expect substantial gains from a well-developed manufacturing sector integrated into international supply chains, while the second may gain from increased productivity of large companies in the wholesale and retail trade sector which have a relatively high share in its economy. Among all the European countries, **Luxembourg** has the highest data-driven growth potential which comes from its developed financial sector. **Ireland** can see comparable economic gains which are explained by the combination of high-tech manufacturing and a relatively well-developed financial sector. Among other European economies which may likely lead the data-driven economic growth, **Sweden** is an example of a country which will highly depend on the data-based innovations in manufacturing while the **United Kingdom** represents an economy heavily reliant on advances in the financial sector. Overall, while different countries may rely on different branches, the broad effects of big and open data mean that total gains will mainly depend on the broad ICT adoption rate and the size and structure of enterprises.

Countries with larger, more global enterprises as well as those more advanced in the ICT field, should benefit more than those that already lag behind in these areas.



Therefore one can say that, big and open data poses different challenges for different parts of the EU. The likely winners will be the North European countries which have favourable structural compositions, both in terms of sectors constituting their economies and the average company size. A major challenge for them will be to remain globally competitive and reshape their industries through investments in the data-driven innovative solutions. On the other hand, the New Member States can also expect substantial gains based on an increasing efficiency but in their case limited investments in the knowledge-intensive services, R&D and innovations may become a serious bottleneck. Reaping the benefits of the data deluge seems to be the biggest challenge for the South European states – both sectoral composition and the average company size here are less compatible with big data applications than in the Northern Europe. While shifts in the economic structure may be difficult in the short run, these countries can increase their potential returns from big data investment by focusing on the data-based managerial skills development and stronger linkages between big data providers and SMEs. Regardless of the country, big and open data impact on the economy will be primarily felt by European citizens. Better tailored public and private services and products as well as more transparent public institutions, will result in clear benefits reflected in their annual incomes which, by 2020, should increase by several hundred euros on average, provided that the big and open data revolution materializes in the EU to the full scale.



Big and open data in Europe
North
South
NMS
Key areas
Switch to data-driven decision making
Ensuring that SMEs will have access to big data solutions (e.g. by third parties)
Efficiency improvements in manufacturing and trade
Growth opportunities
Re-igniting industrial growth through data-based innovative solutions
Improving public sector efficiency, fostering economy restructuring
Fostering knowledge-intensive services growth
Challenges
Securing talent for global competition
Achieving scale effects, securing financing for innovative data solutions, closing the ICT gap
Decreasing innovation lag, enacting effective open data policies
Source: WISE Institute with BOUDICA
Additional output in 2020
Total bn €
per capita €
150 North 560
16 NMS 148
40 South 311
206 EU-28 408
Additional increase in GDP by region
%
2013-2020
Business processes efficiency (Big Data)
Increased competition (Open Data)
Data-driven decision making (Big and Open Data)
0
0,5
1,0
1,5
2,0

Country groups
presented
in the chapter

EU-28

North

Belgium, Denmark, Germany, Ireland, France, Luxembourg, The Netherlands, Austria, Finland, Sweden, The United Kingdom

NMS

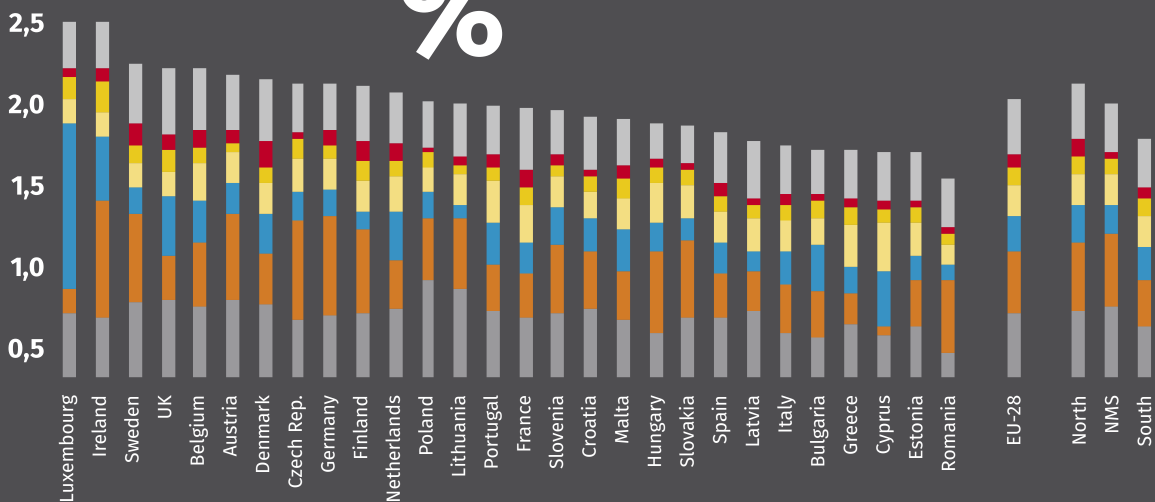
Bulgaria, The Czech Republic, Estonia, Croatia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Romania, Slovenia, Slovakia

South

Italy, Greece, Spain, Portugal

Additional **increase** in GDP level by country and sectors

% 2013-2020



Summ

In order to take full economic advantage of the information and communication technologies, Europe should not overlook the potential of the big and open data revolution that is currently unfolding globally.

In the second half of the 20th century electronics have managed to transform our lives and jobs by supplementing human brains with ever faster machine computational power. Exponential advances in processor speed and capacity of data storage led us to the point where technological ability to register, store and process information is not a real limit any more for human imagination. At the same time, however, inventing methods for economic exploitation of big data still remain a conceptual and managerial, if not technological, challenge. Although nowadays the number of useful business applications of big data is still limited, we can expect that in ten to fifteen years this picture will change dramatically and the winners and losers of this competition will be known.

To avoid losing the global race for data-driven value Europe must, first and foremost, recognize the nature of the big and open data revolution.

The most important aspect from the perspective of European policymakers is the observation that big and open data developments constitute the current phase of a broader ICT revolution, and build upon previous accomplishments in this area. Therefore, strong ICT foundations are needed to fully grasp the economic value of big and open data applications. This concerns the already developed ICT infrastructure, human skills and knowledge acquired in the pre-

ary & recommendations

vious phases of the same digital revolution, but also the ability to innovate in areas which are developing dynamically today: artificial intelligence, large data handling or robotisation. At the same time, a technological dimension of the big and open data change requires complimentary advancement in the management methods before its full potential can be exploited economically. Therefore the entrepreneurship element of the puzzle is of the same importance as the technological one. Policymakers should also remember that the value of data-driven innovations heavily relies on the integration of various data sources and the ability to use and re-use them in a diversified way. Thus, open data complements big data and datafication of the economy in a self-reinforcing process. The more data available for use, the more valuable existing and new data becomes. These positive spillovers and related snow-ball effects provide a sound basis for public action facilitating the development of big data market and opening up of ever more data. Europeans face the triple challenge. Firstly, they must recognize the different initial readiness of the individual Member States for the challenges and opportunities of a data-driven economy. Secondly, they should address them with policy choices tailored to the local specificities, but at the same time strengthening the scale of the common market. Thirdly, they must look on the big and open data challenge not as an isolated regulatory challenge but rather as a part of broader reform agenda for Europe that should reinvent our thinking about the post-crisis economic reality, and provide new engines of economic growth for the future.

Below we present the actions which we propose to be taken by the EU and the Member States in order to ensure capturing the entire economic potential of big and open data.

I . Improving Fundamentals

Strengthening the single market.

Investing in technical and managerial human capital.

Providing European-level support for innovative, high-growth start-ups to make them global.

Successfully seizing this big and open data opportunity will, depend on the broad, structural challenges which currently place Europe in a relatively unfavourable position vis-à-vis its global competitors. The three key areas where action is needed are the single market, human capital and innovative ventures

The advent of big data provides an additional motivation for deepening European economic integration as the scale, variety and freedom of data flows available to enterprises become an important factor for their success. This is especially important for increasing competitiveness of the European ICT sector as well as fostering productivity growth in data-intensive services. Creating a truly single digital and services markets will be as important as ever for the future of the EU economy.It should lower the entry barriers to the big data industry in Europe, enabling an easier creation of universal, pan-EU ICT solutions with lower costs for final consumers and the public sector.

While the shortage of skilled data scientists may be evident, the second important element of data-driven economic growth – managers ready to capture the value of novel data-driven solutions – is often missing from the debate. As data has no impact without context, so will the brightest data analysts have no impact on their enterprises without recognition of the value of implementing data-based solutions. Therefore, investing in the quantitative training in management studies should not be underestimated.

One of the major economic disadvantages of Europe is the lack of successful, innovative ventures, particularly in the area of digital economy. The main barriers to growth of start-ups in the EU are linked to insufficient European integration. Access to high-risk financing for EU start-ups is hindered by the lack of a single market for venture capital, cross-border cooperation between business sector and research institutions is limited and Public sector support for start-ups at the national level is fragmented. To address this problem, coordinated action at the EU level is needed. The regulations concerning high-risk financing should be harmonized to create a single European market, cross-border cooperation between SMEs and research institutions should be promoted (including providing more support to the best institutions while strengthening their pan-European linkages). Overall, both the regulatory environment and funding opportunities for young enterprises in every European country should be sufficient to go European and global from the start, without the need to cope with almost thirty different legal systems.

II . Building innovative potential

Strengthen demand for big data solutions provided by Big Science projects.

Ensure support for research, development and demonstration of ICT innovations.

Provide support for exploring innovative big and open data applications in SMEs.

To gain and maintain a competitive edge in the area of big and open data applications the EU should support the whole innovation chain, from basic research and early experiments with novel ways of analysing data, through linking it to other relevant ICT solutions, to finding new business models applicable to the European economy

While big data technologies have already moved from the labs to the enterprises, large-scale scientific projects are still important engines of data analytics progress. Providing financial and organizational support for Big Science projects is as important as ever, as analytical techniques developed during research may prove as valuable for the economy as its core findings.

Big and open data projects rely on a wide range of enabling technologies which allow for the collection, transmission, storage and processing data. Without broader ICT solutions such as cloud computing or wirelessly connected sensors the potential value of data will be stuck in a technological bottleneck.

SME engagement in data-driven innovation is crucial for European economy because large companies constitute only a fraction of enterprises and are responsible for less than 40% of European GDP. Finding novel ways of providing affordable access to data-driven innovations for the SMEs (e.g. through Data- or Analytics-as-a-Service business models) should be encouraged. One possible source of funding for such projects is COSME.

III . Stimulate uptake by soft measures

Raise awareness of potential open data uses and business models based on open data.

Public administration should lead by example – experiment with assessing policies and managing institutions, follow best practices.

Providing support for big and open data uptake in public administration as part of the structural policies, especially in the NMS.

Many uses of big and open data, such as the adoption of Data-to-Management solutions or finding novel uses for freely available datasets require not only appropriate infrastructure or skills, but also awareness of new business opportunities. Public authorities may use soft measures to promote the uptake of data-driven innovations.

Stimulating interest for open data increases chances that its potential value will be fully uncovered by innovative individuals and enterprises. Enlarging the pool of available data by promoting open data business models provides positive spillovers to the whole economy as additional valuable discoveries can be made by combining and re-using the expanding variety of datasets.

Initiating the change with themselves, public authorities will not only enhance their own efficiency, but also provide market pull for innovative data-based solutions.

Research shows varying levels of engagement in open data projects around the EU. The uptake of the ICT solutions and data-based solutions may be a challenging process for public administration. Building the capacity of authorities to realize the potential of big and open data provides benefits for both public and private sectors and fosters cohesion in the EU by bridging gaps in the performance of public administration.

IV . Enlarge the data pool

while maintaining data holders' interests

Harmonizing regulations and standards of data on public services (including health and education).

Consider the approach to personal data protection.

Improve economic data collection and use it to assess economic policies and provide insights to the private sector.

In many cases the viability of extracting value from data depends on regulations. The EU should continue to harmonize data regulations and standards across the continent in order to balance interests of data holders and users without creating unnecessary legal burdens.

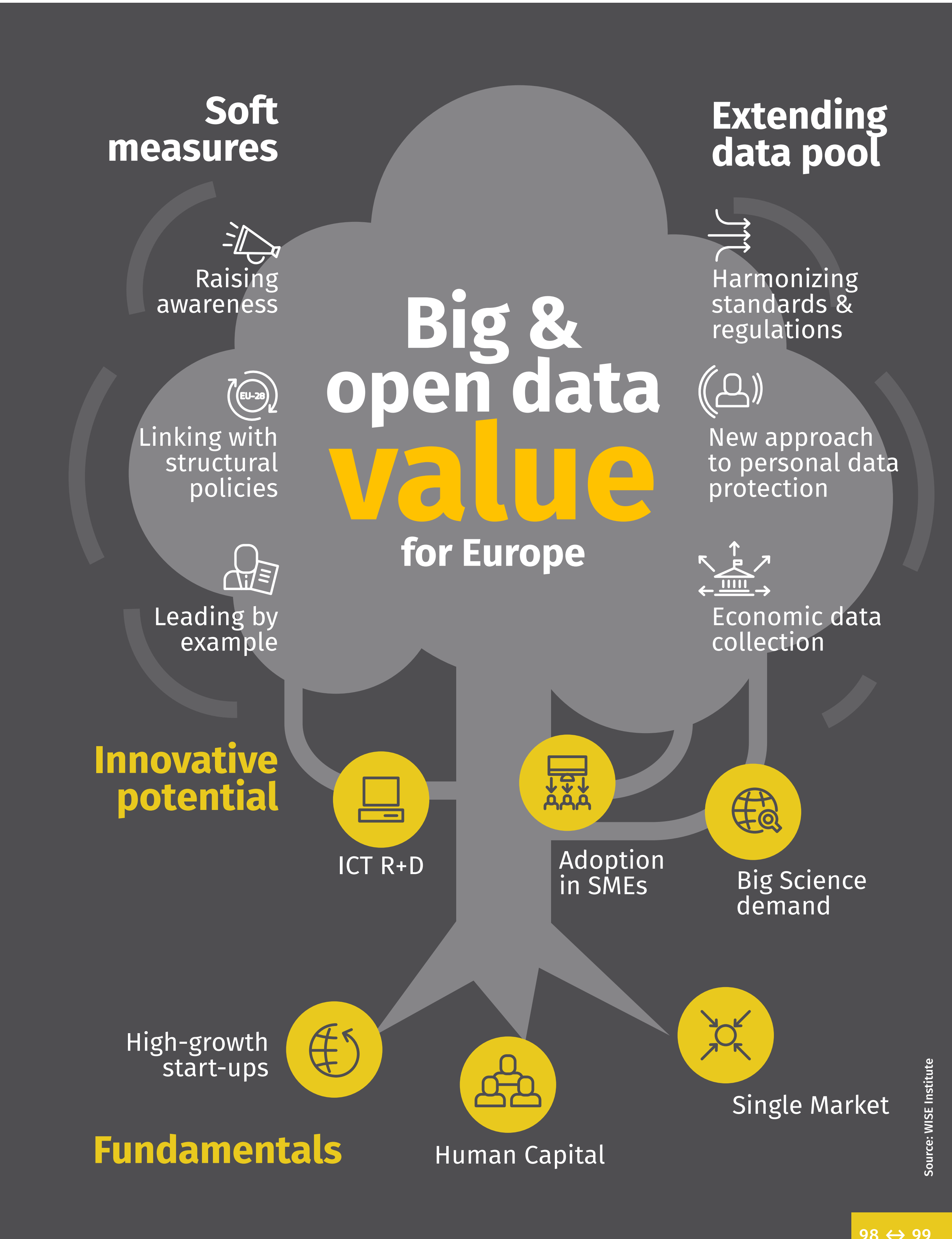
The data-driven innovations offer new ways for improving public services. To ensure better comparability and thus better insights gained from data on public services from different European countries, it is necessary to provide a common regulatory framework. Developing common standards for health data may be especially important in the context of the looming demographical and economic challenges of the ageing European population. The potential of ICT innovations to improve the efficiency of the European healthcare must be used to its fullest. Additional emphasis should be put on ensuring that verification of the public open data authenticity and integrity (e.g. through electronic signatures).

While protecting the right to privacy is of utmost importance, the rules of personal data protection could also be taken into account. The possibility of simplifying user agreements should be investigated. Specifying what ways of data use and re-use are not allowed under different broad options of personal data protection in the legislation and then letting the individual choose one of them may provide both protection of individual's interests and freedom for the private sector to look for value in the acquired data under clear and stable rules.

As the financial flows become ever more digitized, the main challenges of detailed monitoring of the economy are no longer technical, but rather legislative. Public authorities should explore ways in which the dispersed digital data on the functioning of the economy may be aggregated without creating privacy, confidentiality and national security risks. The gains from closing information gaps on the real picture of the economy should be great in both the public (better policies) and the private (better business decisions) sector.

Big and open data brings new, smart growth opportunities for Europe.

As data becomes an essential resource for the modern economy, European countries get yet another chance to re-ignite their stalling economic engines. European governments, enterprises, research institutions and individuals have this new resource at hand. Whether they will fully use it to their advantage depends entirely on Europe's own innovativeness and ability to set up a coherent framework to explore the hidden value of data. While resource constraints are not an issue for big and open data, both old and new institutional challenges may pose a threat to successful implementation of this new wave of ICT innovations. A partially functioning internal market, human capital shortages, institutional inertia and conflicts over balancing freedom of data use with data protection all need addressing in the near future if Europe wants to reap the full benefits of big and open data by the end of the decade.


Soft measures
Raising awareness
EU-28
Linking with structural policies
Leading by example
Big & open data value for Europe
Extending data pool
Harmonizing standards & regulations
New approach to personal data protection
Economic data collection
Innovative potential
ICT R+D
Adoption in SMEs
Big Science demand
High-growth start-ups
Human Capital
Single Market
Fundamentals
Source: WISE Institute

Litera

ture

A

Aghion P., Howitt P.W. (1998), **Endogenous Growth Theory**; MIT Press, Cambridge.

Ark B. van, O'Mahony M., Timmer M.P. (2008), **The Productivity Gap between Europe and the United States: Trends and Causes**; Journal of Economic Perspectives 22 (1), 2544.

Autor D.H, Levy F., Murnane R.J. (2003), **The Skill Content Of Recent Technological Change: An Empirical Exploration**; The Quarterly Journal of Economics, 118(4), 1279-1333, MIT Press, Cambridge.

B

Berndt E.R., Morrison C.J. (1995), **High-tech capital formation and economic performance in U.S. manufacturing industries: An exploratory analysis**; Journal of Econometrics 65 (1), 943.

Black S.E., Lynch L.M. (2001), **How to Compete: The Impact of Workplace Practices and Information Technology on Productivity**; The Review of Economics and Statistics 83 (3), 434-445.

Black S.E., Lynch L.M. (2002), **What's Driving the New Economy? The Benefits of Workplace Innovation**; Center for Economic Studies Working Papers 02-03, U.S. Census Bureau.

Bloom N., Sadun R., van Reenen J. (2010), **Americans do I.T. Better: US Multinationals and the Productivity Miracle**; NBER Working Papers 13085, National Bureau of Economic Research.

Bosworth B.P, Triplett J.E. (2003), **Services Productivity in the United States: Griliches' Services Volume Revisited**; manuscript.

Bresnahan T.F., Brynjolfsson E., Hitt L.M. (2002), **Information Technology, Workplace Organization, And The Demand For Skilled Labor: Firm-Level Evidence**; The Quarterly Journal of Economics 117 (1), 339-376.

Bresnahan T.F., Trajtenberg M. (1992), **General Purpose Technologies: Engines of Growth**, Papers 16-92; Tel-Aviv.

Bresnahan T.F., Trajtenberg M. (1995), **General purpose technologies 'Engines of growth'?**; Journal of Econometrics 65 (1), 83-103; Elsevier.

Brynjolfsson E. (1994), **Some Estimates of the Contribution of Information Technology to Consumer Welfare**; MIT Center for Coordination WP 161.

Brynjolfsson E., Hitt L.M., Kim H.H. (2011), **Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance, available at SSRN** (http://ssrn.com/abstract=1819486).

Brynjolfsson E., Kemerer C.F. (1996), **Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market;** Management Science 42(12), 1627-1647.

C

Carlsson B. (2004), **The Digital Economy: what is new and what is not?**; Structural Change and Economic Dynamics 15, 245-264.

Cavallo A. (2011), **Scraped Data and Sticky Prices**, Working Paper, MIT Sloan.

Cebr (2012), **Data equity. Unlocking the value of big data**, report for SAS, Centre for Economics and Business Research.

Chennells L., van Reenen J. (1999), **Has technology hurt less skilled workers? A survey of the micro-econometric evidence**; Institute for Fiscal Studies WP W99/27.

Chesbrough H.W., Teece D.J. (2002), **Organizing for Innovation: When is Virtual Virtuous?**; (HBR Classic), Harvard Business Review, August, 127-134.

Clarke L., Weyant J., Birky A. (2006), **On the sources of technological change: Assessing the evidence**; Energy Economics 28 (5-6), Elsevier.

Crespi G., Criscuolo C., Haskel J. (2007), **Information Technology, Organizational Change and Productivity Growth: Evidence from UK Firms**; 558.

Cukier K., Schönberger V. (2013), **Big data: a revolution that will transform how we live, work, and think**, Boston: Houghton Mifflin Harcourt.

D

Daveri F. (2002), **The New Economy in Europe**, 1992 - 2001; Oxford Review of Economic Policy 18 (3), 345-362.

David P.A. (2001), **Digital Technologies, Research Collaborations and the Extension of Protection for Intellectual Property in Science: Will Building ‚Good Fences' Really Make, Good Neighbors'?**; MERIT Research Memoranda 2001-004.

David P.A. (2010), **The Dynamo and the Computer: An Historical Perspective on the Productivity Paradox**; Levine's WP Archive 115, David K. Levine.

David P.A., Wright G. (1999), **General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution**; manuscript.

David P.A., Wright G. (2005), **General Purpose Technologies and Productivity Surges: Historical Reflections on the Future of the ICT Revolution**; Economic History, EconWPA.

David, P.A. (1991), **Computer and Dynamo: The Modern Productivity Paradox in a Not-too-distant Mirror, in Technology and Productivity;** OECD, Paris.

Deloitte (2013), **Open growth. Stimulating demand for open data in the UK, briefing note,** Deloitte Analytics.

Dewan S., Kraemer K.L. (2000), **Information Technology and Productivity: Evidence from Country – Level Data;** Management Science 46 (4), 548-562.

EC (2009), **Internet of Things — An action plan for Europe**, COM(2009) 278 final, European Commission, Brussels.

E

Edquist H. and Henrekson M., (2006), **Technological Breakthroughs and Productivity Growth**, Working Paper Series 665, Research Institute of Industrial Economics.

Einav L. Levin J.D. (2013), **The Data Revolution and Economic Analysis**, NBER Working Papers 19035, National Bureau of Economic Research.

ePSI scoreboard, epsiplatform.eu/content/european-psi-scoreboard, European Public Sector Information Platform.

Etro F. (2009), **The Economic Impact of Cloud Computing on Business Creation, Employment and Output in Europe. An application of the Endogenous Market Structures Approach to a GPT innovation**, Review of Business and Economics, Katholieke Universiteit Leuven, Faculteit Economie en Bedrijfswetenschappen, vol. 0(2), 179-208.

Evans D., **The Internet of Things. How the Next Evolution of the Internet Is Changing Everything**, White Paper, CISCO IBSG.

Evans P.C., Annunziata M. (2012), **Industrial Internet: Pushing the Boundaries of Minds and Machines**, report, GE.

G

G8 (2013), **G8 Open Data Charter**, the Group of Eight, Lough Erne.

Gordon R.J. (2003), **Exploding Productivity Growth: Context, Causes, and Implications**; Brookings Papers on Economic Activity 2003 (2), 207-279.

Grace K. (2013), **Algorithmic Progress in Six Domains**. Technical Report, 2013-3, Machine Intelligence Research Institute, Berkeley, CA.

Granstrand O., Patel P., Pavitt K. (1997), **Multi-Technology Corporations: Why They Have "Distributed" Rather Than "Distinctive Core" Competencies'**; California Management Review 39 (4), 8-25.

Griliches Z. (1992), **The Search for R&D Spillovers**; Scandinavian Journal of Economics 94 (0), 29-47.

Gruber H., Koutroumpis P. (2010), **Mobile Telecommunications and the Impact on Economic Development**; manuscript.

Guerrieri P., Padoan P.C. (eds.) (2007), **Modelling ICT as a General Purpose Technologies**; Collegium 35.

H

Hayek F.A. von (1974); **The Pretence of Knowledge**, Nobel Prize in Economics documents 1974-2, Nobel Prize Committee.

Hempell T. (2005), **What's Spurious, What's Real? Measuring the Productivity Impacts of ICT at the Firm–Level**; Empirical Economics 30 (2), 427-464.

Hempell T. (2006), **Computer and Productivity: How Firms Make a General Purpose Technology Work**; ZEW Studies 33.

Hilbert M., López P. (2011), **The World's Technological Capacity to Store, Communicate, and Compute Information**, Science vol. 332 (6025), 60-65.

I

IBM (2012), http://ibmresearchnews.blogspot.com/2013/03/dimitri-kanevsky-translating-big-data.html

IERC (2012), **The Internet of Things 2012. New Horizons**, Internet of Things European Research Cluster, Halifax.

Inklaar R., Timmer M.P., van Ark B. (2008), **Market services productivity across Europe and the US**; Economic Policy 23 (53), 139-194.

J

Jorgenson D.W., Ho M.S., Stiroh K.J. (2008), **A Retrospective Look at the U.S. Productivity Growth Resurgence**; Journal of Economic Perspectives 22 (1), 324.

Jorgenson D.W., Stiroh K.J. (1999), **Information Technology and Growth**. American Economic Review 89 (2), 109-115.

Jovanovic B., Rousseau P.L. (2003), **Two Technological Revolutions**, Journal of the European Economic Association, MIT Press, vol. 1(2-3), 419-428.

Jovanovic B., Rousseau P.L. (2005), **General Purpose Technologies**; Handbook of Economic Growth 1B, 1181-1224; Elsevier.

K

Kretschmer T. (2012), **Information and Communication Technologies and Productivity Growth: A Survey of the Literature**, OECD Digital Economy Papers No. 195, OECD Publishing.

Walter C. (2005), **Kryder's Law**, Scientific American 06/2005

L

Lipsey R.G., Bekar C., Carlaw K. (1998), **What requires explanation?** [in:] Helpman E. (edt.); General purpose technology and Economic Growth; MIT press, Cambridge.

Lipsey R.G., Carlaw K.I., Bekar C.T. (2005), **Economic Transformations. General Purpose Technologies and Long Term Economic Growth**; Oxford University Press, Oxford.

Lucas R. Jr (1976); E**conometric policy evaluation: A critique**, Carnegie-Rochester Conference Series on Public Policy, Elsevier, vol. 1(1), 19-46, January.

M

Marrano M.G., Haskel J., Wallis G. (2009), **What happened to the knowledge economy? ICT, Intangible Investmet, and Britain's Productivity Record Revisited**; Review of Income and Wealth 55 (3), 686-716.

MasterCard Advisors (2013), **Cashless Journey**, online report.

McKinsey Global Institute (2011), **Big data: The next frontier for innovation, competition, and productivity**, Washington, D.C.

Milgrom P., Qian Y., Roberts J. (1991), **Complementarities, Momentum, and the Evolution of Modern Manufacturing**; American Economic Review 81 (2), 84-88.

Moore, G.E. (1965), **Cramming More Components onto Integrated Circuits**, Electronics, 114–117, April 19.

O

OECD (2013); **Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by „Big Data"**, OECD Digital Economy Papers, No. 222, OECD Publishing.

Oliner S.D., Sichel D.E., Stiroh K.J. (2007), **Explaining a Productive Decade**; Brookings Papers on Economic Activity (1), 81-137.

Open Definition, opendefinition.org, Open Knowledge Foundation.

Oz E. (2005), **Information technology productivity: in search of a definite observation**; Information & Management 42 (6), 789-798.

R

Rosenberg N. (1998), **Chemical Engineering as a General Purpose Technology**; [in:] Helpman E. (edt.); General purpose technology and Economic Growth; MIT press, Cambridge.

Rosenberg N., Trajtenberg M. (2004), **A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States**; The Journal of Economic History 64 (01) 61-99, Cambridge.

S

Schreyer P. (2001), **Information And Communication Technology And The Measurement Of Volume Output And Final Demand – A Five-Country Study**; Economics of Innovation and New Technology 10 (5), 339-376.

Shiu A., Lam P.L. (2008), **Causal Relationship between Telecommunications and Economic Growth in China and its Regions**; Regional Studies 42 (5), pp. 705-718.

Spitz-Oener A. (2003), **IT Capital, Job Content and Educational Attainment**; ZEW DP 03-04; Centre for European Economic Research (ZEW), Mannheim.

Stiroh K.J. (2002), **Are ICT Spillovers Driving the New Economy?**; Review of Income and Wealth 48 (1), 33-57.

Stranacher K., Krnjic V., Zefferer T. (2013), T**rust and Reliability for Public Sector Data**, World Academy of Science, Engineering and Technology 73, pp. 384-390.

T

Tambe P., Hitt L.M. (2010), **Job Hopping, Information Technology Spillovers, and Productivity Growth**, SSRN eLibrary 1302637.

Timmer M.P., Ypma G., van Ark B. (2003), **IT in the European Union: Driving Productivity Divergence?** Research Memorandum Groningen Growth and Development Centre 67.

Trajtenberg M., Yitzhaki S. (1982), **The Diffusion of Innovations: A Methodological Reappraisal**; National Bureau of Economic Research WP 1008.

U

UN Global Pulse (2012), **Big Data for Development: Challenges & Opportunities**, White Paper, United Nations.

W

White House (2013), **Open Data Policy – Managing Information as an Asset**, Memorandum for the Heads of Executive Departments and Agencies, Office of Management and Budget, Executive Office of the President, Washington D.C.

Appendix A:

Big and Open Data Universal Impact Assessment Model (BOUDICA)

Model structure

The bottom-up BOUDICA projection model (**B**ig and **O**pen **D**ata **U**niversal **I**mpa**c**t **A**ssessment model) provides estimates of big and open data impact on enterprises in various sectors as this affects their ability to pursue big data projects. BOUDICA incorporates detailed data on the European economy extracted from the Eurostat database. The starting year for all calculations is 2010, as this is the last year represented there for the whole cross-country sample. Results are presented for years 2013–2020.

Total impact of big and open data on the EU-28 GDP in a given year of analysis can be encapsulated by the following formula:

$$\Delta GDP_{EU28}(t) = \sum_c \sum_n \sum_s [(I_{cns}^{DDD}(t) + I_{cns}^{OD}(t) + I_{cns}^{BD}(t)) \times VA_{csn}],$$

where

ΔGDP_{EU28} – impact of big and open data on the EU-28 GDP,

t – time variable, t = *analysed year – base year (2010)*,

c – country index,

n – NACE sectors index,

s – enterprise size index,

VA_{csn} – total value added by enterprises of size s operating in the sector n in country c in the base year (2010),

$I_{cns}^{DDD}(t)$ – total impact of data-driven management on enterprises of size s operating in sector n in country c after t years,

$I_{cns}^{OD}(t)$ – total impact of increased competition on enterprises of size s operating in sector n in country c after t years,

$I_{cns}^{BD}(t)$ – total impact of efficiency improvements on enterprises of size s operating in sector n in country c after t years.

The formula may be easily modified to obtain more detailed estimates by country, sector and company size.

Modelling the big and open data impact

The three main types of big and open data impacts are assumed to gradually spread across the economy as more enterprises start to exploit created by them and as productivity growth on microeconomic level. The impacts are calculated as follows:

$$I(t)_{cns}^{DDD} = a_{cns}^{DDD}(t) \times i^{DDD}(1) + a_{cns}^{DDD}(t-1) \times i^{DDD}(2) + \ldots + a_{cns}^{DDD}(1) \times i^{DDD}(t),$$

$$I(t)_{cns}^{OD} = a_{cns}^{OD}(t) \times i^{OD}(1) + a_{cns}^{OD}(t-1) \times i^{OD}(2) + \ldots + a_{cns}^{OD}(1) \times i^{OD}(t),$$

$$I(t)_{cns}^{OD} = a_{cns}^{BD}(t) \times i_{n}^{BD}(1) + a_{cns}^{BD}(t-1) \times i_{n}^{BD}(2) + \ldots + a_{cns}^{OD}(1) \times i_{n}^{BD}(t),$$

where $i^{DDD}(t)$, $i^{OD}(t)$ and $i_{n}^{BD}(t)$ denote impact parameters based on literature and fully realizing during the several years after the effect affects the group of enterprises for the first time (hence it is the function of time), while $a_{cns}^{DDD}(t)$, $a_{cns}^{OD}(t)$ and $a_{cns}^{BD}(t)$ denote the percentage of enterprises affected by a given effect in a given year. The latter values are calculated from the adoption rate paths ($A_{cns}^{DDD}(t)$, $A_{cns}^{OD}(t)$ and $A_{cns}^{BD}(t)$) modelled for enterprises of different sizes and operating in different sectors and countries.

$$a_{cns}^{DDD}(t) = A(t)_{cns}^{DDD} - A(t-1)_{cns}^{DDD},$$

$$a_{cns}^{OD}(t) = A(t)_{cns}^{OD} - A(t-1)_{cns}^{OD},$$

$$a_{cns}^{BD}(t) = A(t)_{cns}^{BD} - A(t-1)_{cns}^{BD}.$$

Impact of big and open data adoption – key parameters

Type of effect	Impact	Source
Data-driven decision making	5% productivity growth in all sectors	Brynjolfsson et al (2011)
Open data decreasing barriers to entry	0.2% growth in value added (impact varies by sectors)	Own estimates based on Etro (2009) study on effects of introducing cloud computing
Efficiency improvements – supply chain management	10% productivity growth in retail and wholesale trade	Own estimates based on McKinsey Global Institute (2011) study
Efficiency improvements – industrial processes	10% productivity growth in manufacturing	Own estimates based on McKinsey Global Institute (2011) study
Efficiency improvements – fraud/error detection	2.5% productivity growth in financial industry and public administration	Own estimates based on McKinsey Global Institute (2011) and Cebr (2012) studies

Note: Own estimates take into account applicability of the bottom-up analysis results to the macro level and robustness of underlying assumptions.

Modelling the adoption rate paths

The annual adoption rates were modelled on country-by-country and sector-by-sector basis taking into account company size. The dynamics of adoption $A(t)$ were modelled as an S-function ($S(t)$) estimated on the survey data about the adoption of previous ICT innovations by the enterprises (homepages, DSL, fixed broadband access, LAN and Intranet/extranet, online banking and financial services usage) and then scaled to account for different saturation levels across countries, sectors and companies of different size.

$$A(t)^{DDD}_{cns} = S(t) \times SF^{DDD}_{cns}$$

$$A(t)^{OD}_{cns} = S(t) \times SF^{OD}_{cns}$$

$$A(t)^{BD}_{cns} = S(t)^{BD}_{cns} \times SF^{BD}_{cns}$$

The scaling factors (SF) for each country were based on a combination of general ICT index and an additional impact-specific index based on relevant literature. For data-driven management practices and efficiency improvements related to big data, the additional index (BD_{cns}) was based on McKinsey Global Institute (2011) and Cebr (2012) assessments of sectoral readiness for big data solutions. For data-driven decision making the index was corrected upwards for health and education services to take into account additional motivational effects of open data from the public sector. For increased competition stemming from opening access to data, the additional index (OD_{cns}) was based on results of Deloitte (2013) study (sectoral differences) and ePSI Scoreboard (country-level differences).

$$SF^{DDD}_{cns} = SF^{BD}_{cns} = ⅓ \times ICT_{cns} + ⅔ \times BD_{ns}$$

$$SF^{OD}_{cns} = ½ \times ICT_{cns} + ½ \times OD_{cn}$$

ICT Index

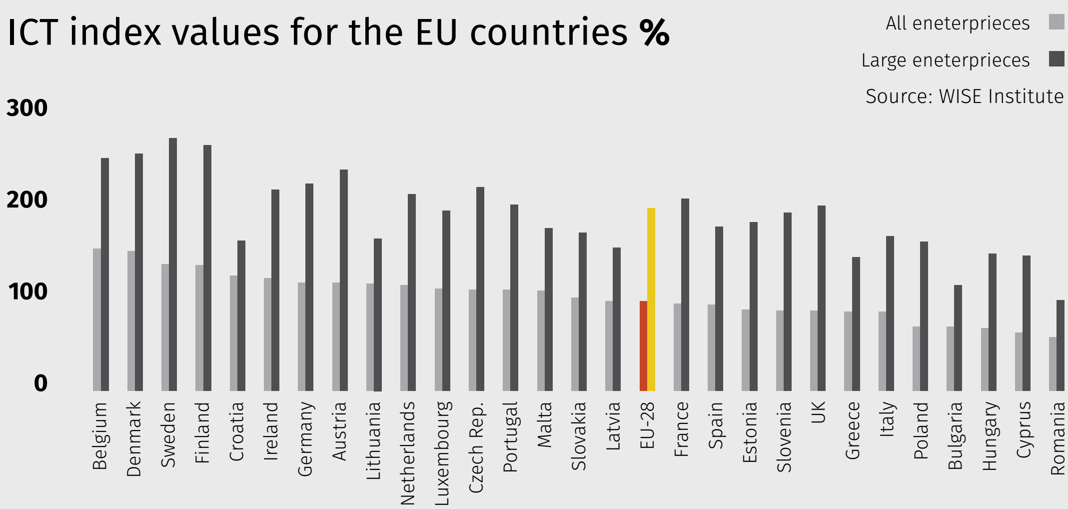
The ICT index was calculated by combining data from the surveys of ICT adoption by the European enterprises provided by the Eurostat. Indicator values are averages of available data from 2010-2012. They are standardized so that average for all enterprises in the EU is 100% for every indicator.

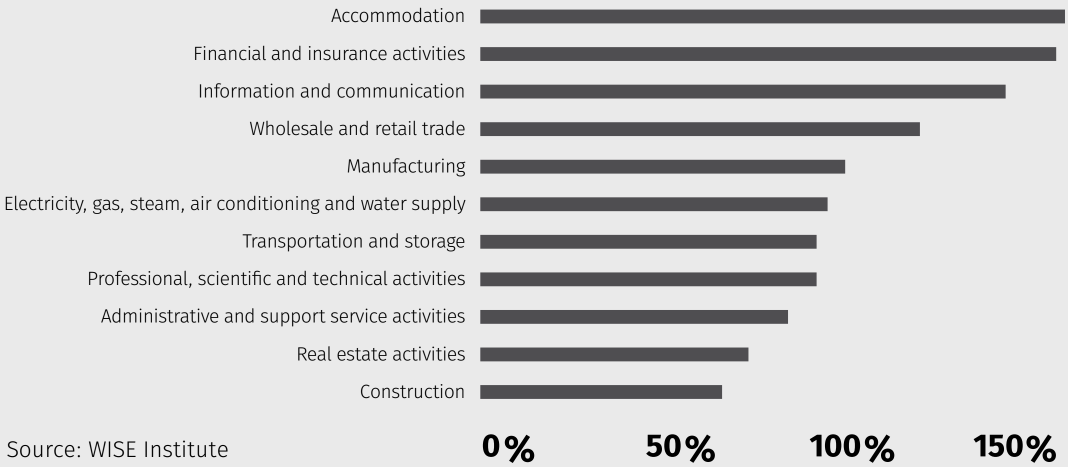
Category	Eurostat Indicator	Weight
Online business activity	Enterprises purchasing online (at least 1% of orders)	10%
	Enterprises selling online (at least 1% of turnover)	10%
Internal infrastructure	Enterprises where persons employed have access to personal human resources services electronically	10%
	Enterprises which electronically share information on purchases with the software used for any internal function	10%
	Enterprises whichelectronically share information on sales with the software used for any internal function	10%
Data exchange	Enterprises which electronically share information suitable for automatic processing with external business partners or on the SCM with suppliers or customers	10%
	Enterprises sending / receiving e-invoices in a standard structure suitable for automatic processing	10%
	Enterprises using automated data exchange with other ICT systems outside the own enterprise	10%
Data collection	Enterprises using Radio Frequency Identification (RFID) technologies	20%

Source: WISE Institute

The ICT index covers all countries and sizes of enterprises as well as most NACE sectors. The missing sectors are estimated based on average values of index for the whole country. ICT index values are higher in large enterprises, especially in the Northern Europe. Sectoral values also differ, indicating that some sectors tend to adopt ICT innovations and invest in appropriate infrastructure significantly more than others, and thus are better prepared for reaping the benefits associated with big and open data.

ICT index values for the EU countries %





Source: WISE Institute

ICT index values for industry and services **%**

Sectors represented in the BOUDICA model

Sectoral breakdown is based on NACE Rev. 2.0 and takes into account 21 sectors.

Accommodation and food service activities	Education	Other service activities
Activities of extraterritorial organisations and bodies	Electricity, gas, steam and air conditioning supply	Professional, scientific and technical activities
Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	Financial and insurance activities	Public administration and defence; compulsory social security
Administrative and support service activities	Human health and social work activities	Real estate activities
Agriculture, forestry and fishing	Information and communication	Transportation and storage

Arts, entertainment and recreation	Manufacturing	Water supply; sewerage, waste management and remediation activities
Construction	Mining and quarrying	Wholesale and retail trade; repair of motor vehicles and motorcycles

Countries analyzed in the BOUDICA model

For presentational purposes, European countries are grouped into three groups. However, big and open data implementation effects are computed for each country separately.

Northern Europe	Belgium, Denmark, Germany, Ireland, France, Luxembourg, The Netherlands, Austria, Finland, Sweden, The United Kingdom
Southern Europe	Italy, Greece, Spain, Portugal
New Member States	Bulgaria,TheCzech Republic, Estonia, Croatia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Romania, Slovenia, Slovakia

Groups of enterprises by size in the BOUDICA model

To calculate the potential adoption rates of innovative data-based solutions, the BOUDICA model takes into account not only the differences among sectors in the given country, but also among companies operating on different scales. There are three groups of companies by size.

Large enterprises	250 or more persons employed
Medium enterprises	50 to 249 persons employed
Small enterprises	fewer than 50 persons employed

Data sources

Sectoral output data was taken from the Structural Business Statistics and National Accounts Detailed Breakdowns in the Eurostat database.

Survey data used for constructing the ICT Index was taken from the Information Society Statistics from the Eurostat database.

Inputs for the open data indices were based on figures from the Deloitte (2013) study and ePSI Scoreboard from the ePSI Platform portal (accessed September 2013). Inputs for the big data indices were based on assessments by McKinsey Global Institute (2011) and Cebr (2012).

ISBN: 978-83-925542-1-9

Project Editor: Krzysztof Blusz
Scientific Editor: Maciej Bukowski, PhD
Cover and design: Bakalie Branding Studio
Printed and bound by: Mdruk

First published in Poland, Warsaw 2014 by:

demosEUROPA
– Centre for European Strategy Foundation
23 Mokotowska Street Apt. 8
00-560 Warsaw, Poland
phone: +48 22 401 70 26
demoseuropa@demoseuropa.eu
http: www.demoseuropa.eu

in cooperation with:

Warsaw Institute for Economic Studies (WISE Institute)
Al. Jerozolimskie 99/18
02-001 Warsaw, Poland
phone: +48 22 395 50 11
wise@wise-instituite.org.pl
http: www.wise-institute.org.pl