# *Chapter 6*

## Skills and employment in a data-driven economy

*This chapter discusses the implications of data-driven innovation (DDI) on skills and employment, focusing on two challenges in particular: one, DDI may further increase pressure on the labour market, and especially on middle income jobs, as it enables an increasing number of cognitive and manual tasks to be performed by data- and analytics-empowered applications; and two, the demand for data specialist skills may exceed supply on the labour market. The chapter first shows that DDI could lead to structural change in labour markets, and discusses the implications with regard to skills. It then focuses on data specialist skills and competence, the lack of which could prevent economy-wide adoption of DDI and the (re-)creation of jobs. Finally, the chapter discusses the policy challenges for promoting DDI while smoothing structural adjustments, focusing on challenges in i) addressing wage and income inequalities, and ii) satisfying skills and competence needs.*

*It's in Apple's DNA that technology alone is not enough – it's technology married with liberal arts, married with the humanities, that yields us the results that make our heart sing.* (Steve Jobs during the launch of Apple's iPad 2 in March 2011)

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The analysis of large volumes of (digital) data, now commonly referred to as "big data", is driving knowledge and value creation across society; fostering new products, processes and markets; and spurring entirely new business models (i.e. data-driven innovation, DDI). Algorithmic trading systems (ATS), for example, analyse massive amounts of market data on a millisecond basis to autonomously identify what to stock and when, and at what price to trade (see Chapter 3 of this volume). ATS, a process unheard of a decade ago, now accounts for more than half of all financial market trading in the United States, and almost a third of all financial market trades in Europe (see Figure 3.12 in Chapter 3).

DDI has the potential to disrupt and transform even traditional sectors such as retail, manufacturing, and agriculture, and thereby to boost economic competitiveness and productivity growth across the economy. Some companies in these sectors are taking advantage of DDI as they are becoming more and more service-like, a trend that some have described using the term "servicification" (Lodefalk, 2010). In manufacturing, for instance, companies are increasingly using sensors mounted on production machines and products, taking advantage of the Internet of Things (IoT). This trend, enabled by machine-to-machine communication (M2M) and analysis of sensor data, has been described by some as the "Industrial Internet" (Bruner, 2013) or "network manufacturing" (Economist Intelligence Unit, 2014). Sensor data are used here to monitor and optimise machine operations at a system-wide level, and for after-sale services, including preventive maintenance operations. DDI is thereby enabling a new generation of highly automated factories, such as the Philips shaver factory in Drachten, the Netherlands, which employs only one-tenth of the workforce it employs in its factory in People's Republic of China (hereafter 'China') that makes the same shavers (see Markoff, 2012).

There is little evidence on the effects of DDI, but the few studies available suggest that firms using DDI raise labour productivity faster than non-users. A study of 330 companies in the United States by Brynjolfsson, Hitt and Kim (2011) estimates that the output and productivity of firms that adopt data-driven decision making are 5% to 6% higher than would be expected from their other investments in, and use of, ICTs. These firms also perform better in terms of asset utilisation, return on equity and market value.

A similar study based on 500 firms in the United Kingdom by Bakhshi, Bravo-Biosca and Mateos-Garcia (2014) finds that businesses that make greater use of online customer and consumer data are 8% to 13% more productive as a result. That study is based on a survey by Bakhshi and Mateos-Garcia (2012), but extended by "matching survey responses about data activities with historical performance measures taken from respondents' company accounts, and by conducting an econometric analysis of the link between business performance and data activity while controlling for other characteristics of the business". The analysis shows that, other things being equal, a one-standard deviation greater use of online data is associated with an 8% higher level of total factor productivity (TFP). Firms in the top quartile of online data use are 13% more productive than those in the bottom quartile. Furthermore, the study shows that "use data analysis" and "reporting of data-driven insights" have the strongest link with productivity growth, "whereas amassing data has little or no effect on its own" (Bakhshi, Bravo-Biosca and Mateos-Garcia, 2014). Another study by Barua et al. (2013) suggests that improving the quality and access to data by 10%, by presenting data more concisely and consistently across platforms and allowing it to be more easily manipulated, would increase labour productivity by 14% on average, but with significant cross-industry variations.

Overall, these studies suggest an approximately 5-10% faster productivity growth compared with non-users. However, it should be stressed that these estimates can hardly be generalised, for a number of reasons. First, as illustrated above, these estimated effects of DDI vary by sector and are subject to complementary factors such as the availability of skills and competences, and the availability and quality (i.e. relevance and timeliness) of the data used. But more importantly, these studies often suffer from selection biases, which make it difficult to disentangle the effects of DDI from other factors at the firm level.[1] More studies are therefore needed to better assess the impact of DDI at that level.

In the current context of weak global recovery, DDI has caught policy makers' attention as a *new source of growth* that can boost the productivity and competitiveness of their economies and industries. However, the disruptive nature of DDI may lead to the "creative destruction" of established businesses and markets, and to a structural shift across the economy, in particular within labour markets. With lingering high unemployment in major advanced economies, however, taking advantage of the process of creative destruction induced by DDI will be particularly challenging, for at least two reasons:

1. DDI may further increase pressure on labour market, in particular on middle income jobs which involve a significant share of tasks that now can be performed by data- and analytics-empowered applications. In particular, DDI enables the automation of an increasing number of cognitive and manual tasks. This includes the use of data analytics for a wider range of intellectually demanding tasks, such as the diagnosis of diseases based on analysis of complex information, including from medical documents. It also involves the use of a new generation of autonomous machines and robots that are no longer restricted to very precisely defined environments, and that can be deployed and redeployed at much faster rates compared to current generation robots.
2. The relatively low availability of the critical data specialist skills and competence required for DDI may prove not just a barrier to the adoption of DDI, but also a missed opportunity for job creation. So far there is little cross-country evidence of a skills shortage or mismatch for DDI. However, some have suggested that the demand for data specialist skills exceeds its supply on the labour market. An Economist Intelligence Unit (2012) survey, for instance, shows that "shortage of skilled people to analyse the data properly" is indicated as the second biggest impediment to make use of data analytics (see also MGI, 2011).[2]

That said, DDI provides huge opportunities for business creation across the data ecosystem for start-ups and small and medium enterprises (SMEs); many of these provide new goods and services that could lead to further job creation opportunities, as discussed in much detail in Chapter 2 of this volume. This chapter, however, does not address these (indirect) job creation opportunities induced by DDI.[3]

## 6.1. "Creative destruction" in labour markets

As highlighted in Chapter 3 of this volume, the ubiquitous deployment of information and communication technologies (ICTs) driven by Moore's Law[4] has led to a number of developments. These include in particular i) sensors interconnected through machine-to-machine communication (M2M) accelerating the "datafication" of the physical world, ii) cloud computing providing practically anyone with super computing power as an utility, and iii) data analytics empowering decision support and automation for almost every application area. The confluence of these developments can be expected to reach its tipping point once M2M bypasses human communication, and the Internet will truly become the Internet of Things (IoT – see Box 6.1).[5] This signals a new phase of DDI that today is still in its infancy even in the most advanced economies, and in which software empowered by data analytics could become a major source for labour productivity growth. As Andreessen (2011) wrote, "software is eating the world", and the world will be served in big chunks of data (TNO, 2013). As data-driven software "is eating the world", labour markets may undergo a more profound structural change than what has been observed so far during the digital revolution.

Box 6.1. **The Internet of Things – A game changer**

One of the main reasons for the sudden breakthrough in smart technologies – like driverless cars or the next generation of robots – is the Internet of Things (IoT), which embeds physical objects in information flows. In the case of driverless cars, for instance, it is the road infrastructure, other cars, and last but not least web services such as online maps that "tell" a car essentially what it needs to know. So it is not necessary to equip a car with e.g. a technical image system as powerful as the image processing systems of humans for it to be able to drive on its own, as was previously assumed.

The power of the human image processing system is so huge that it is very difficult to develop a technical alternative of comparable power, despite the growing abilities of these alternatives (Lee, 2015). But such a system is not necessarily required, because cars can now receive huge amounts of just the right data needed to drive autonomously. In this way, cars can "know" even more about their environment than a human driver. For these very reasons a great number of robotic applications formerly thought impossible will become possible soon. It is not that the sensor systems of the robot are exceptionally good; rather, it is that all devices and machines in the manufacturing plant will give the robot the information it needs. This may include products, related robots in the production line, or external suppliers, so that a dynamic optimisation of the overall production process is possible.

Some have stressed that the IoT is also interconnecting and empowering humans with smart applications, leading to the emergence of an intelligent "superorganism" in which the Internet represents the "global digital nervous system" (Radermacher and Beyers, 2011; O'Reilly, 2014). For 2030, it is estimated that 8 billion people and maybe 25 billion active "smart" devices will be interconnected and interwoven by one single huge information network.[1] The result is the constitution of a gigantic, powerful "superorganism"[2], based on never-ending communication streams.[3]

1. In this context, communication can be seen as one of the most powerful intelligence-enhancing processes we know.

2. Communication is what glues the components of that superorganism together. It has a quadratic growth behaviour with respect to the number of components involved, because communication can take place between each pair of members of the superorganism. In a sense, this observation implies positive network effects.

3. The implications are discussed in detail by Kapitza (2005), who looks at the development and size of human civilisation over the past 3 million years. See also Radermacher and Beyers, 2011 and Solte, 2009.

Source: Herlyn et al., 2015.

### *Starting with the digital revolution and the impact of ICTs*

The question of the effects of DDI on employment follows the broader discussion about the employment impact of technology, and of ICT more specifically. That discussion is linked to the fundamental question about "technological unemployment" that Keynes, almost a century ago, described as follows:

> *We are being afflicted with a new disease of which some readers may not yet have heard the name, but of which they will hear a great deal in the years to come – namely, technological unemployment. This means unemployment due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour. But this is only a temporary phase of maladjustment.* (Keynes, 1930)

The debate on technological unemployment has gained momentum recently in light of current debates over the new potential of automation enabled by data and analytics. Looking at available figures from the Bureau of Labor Statistics on labour productivity and private employment for the United States, scholars such as Brynjolfsson and McAfee (2011) have suggested that the long-run positive relationship between labour productivity growth and employment growth may have been broken since the 1990s. In other words, the labour productivity growth enabled by ICTs in the United States seems not to have led to the creation of further employment.

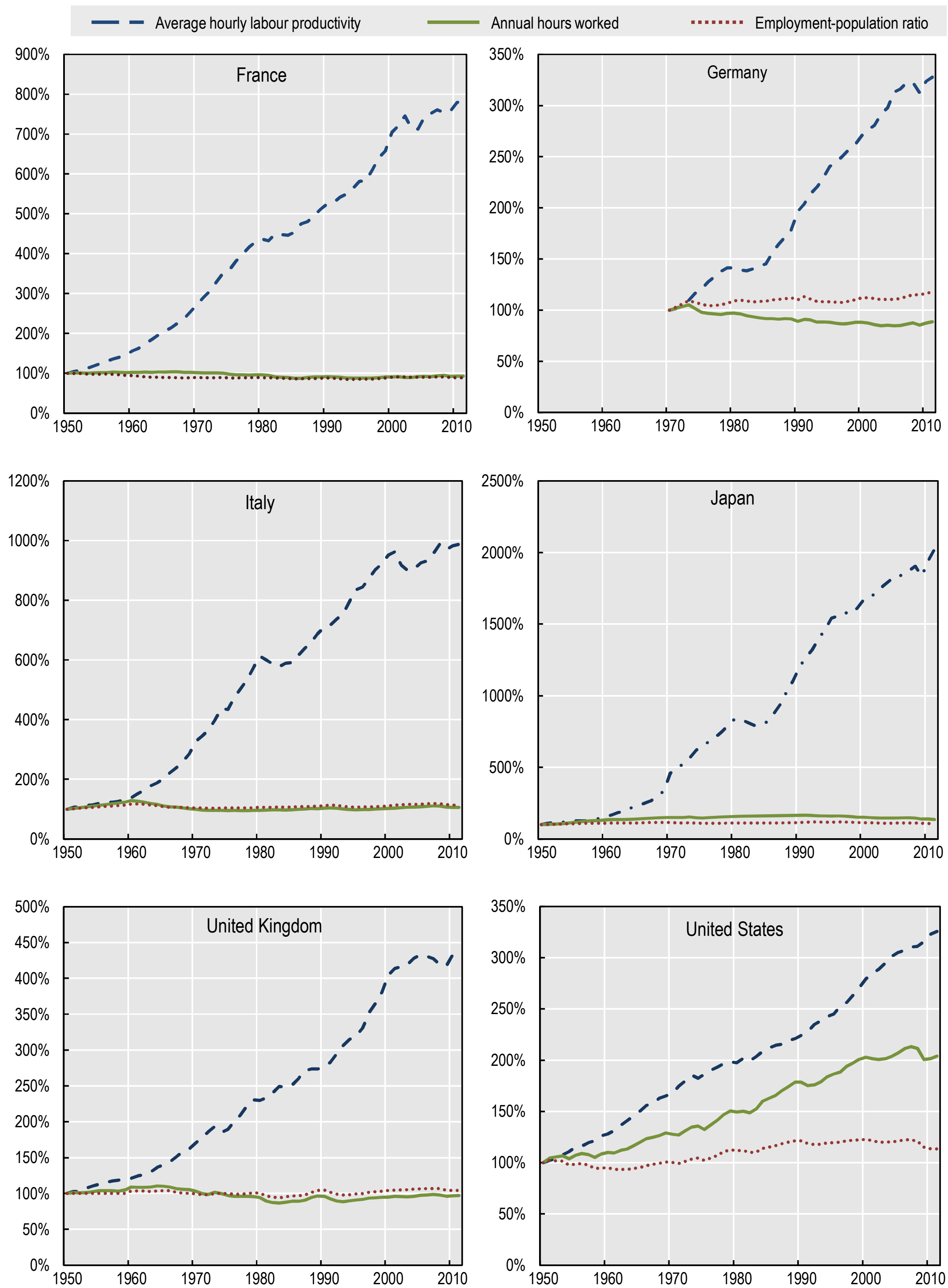
However, such a conclusion is challenged by available cross-country evidence presented in OECD, 2014b, in particular when controlling for the supply of labour. As that report highlights, labour productivity grew at a fast rate in most of the OECD area after the 1960s, while the employment figures in those countries remained stable (Figure 6.**1**). The study therefore concludes that "overall, these long-run trends suggest that compensation mechanisms have been rather effective to maintain employment levels over the last 60 years despite high rates of technological progress" (OECD, 2014b). It should be acknowledged here that this conclusion assumes a "closed" economy where the compensation effects remain within national borders, an assumption challenged by the global and cross-border nature of the data economy. (See the discussion on base erosion and profit shifting, BEPS, in Chapter 2 of this volume.)

Furthermore, many authors – such as Goos, Manning and Salomons (2009) and Autor and Dorn (2013) – have observed a trend of employment polarisation: employment is increasing in both high-skill and low-skill occupations, while stagnating or even declining in middle-skill occupations, with potential negative implications for income equality (OECD, 2014b). "Real wages by skill percentile follow a similar path, suggesting that the increase in employment at the two tails of the skill distribution – high and low skills – has been driven by an increase in demand rather than supply" (OECD, 2014b). According to OECD, 2014b, there are three main explanations for employment polarisation: routinisation, offshoring and international trade. DDI can further leverage routinisation, as it enables and accelerates the automation of some knowledge- and labour-intensive processes.

Furthermore, the increasing use of ICTs across the economy has also driven demand for new types of skills and jobs, most notably linked to ICT specialisation. These are professionals that "have the ability to develop, operate and maintain ICT systems" and for whom "ICTs constitute the main part of their job" (OECD, 2012b). Many of these jobs did not exist before the digital revolution (e.g. software developers), and are rapidly evolving as ICTs progress. In 2013, ICT specialists for the 28 OECD countries considered corresponded to about 14.1 million jobs and to almost 3.5% of total employment, increasing over the decade in all geographic areas (Figure 6.2).

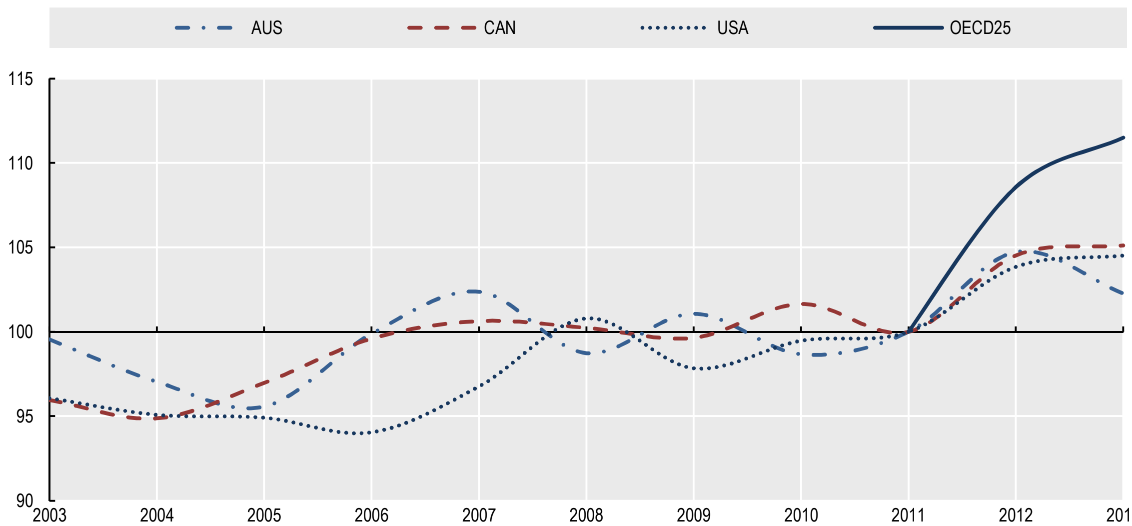
Figure 6.1. **Labour productivity and employment in selected OECD countries (1950-2011)**

1950 (for Germany 1970) = 100%




*Source:* OECD, 2014b based on Penn World Table, version 8.

Figure 6.2. **Trends in the share of ICT specialists in selected OECD countries, 2003-13**
Index 100 = 2011, share in in total employment




*Note:* The OECD25 aggregate includes data for all OECD EU member countries plus Iceland, Norway, Switzerland and Turkey. ICT specialists are defined as ISCO 08 codes: 133, 215, 25, 35 and 742. For Australia, Canada and the United States, correspondence tables were used and some adjustments were needed.

*Source:* OECD based on data from Eurostat, the US Bureau of Census, Statistics Canada, and the Australian Bureau of Statistics labour force surveys.

### *The growing employment polarisation*

Data and analytics have enabled a wide range of "smart" applications that use machine-learning algorithms to "learn" from previous situations, and that can communicate the results of the learning to other machines (see Chapter 3 of this volume).[6] Having analysed similar situations, "smart" applications can infer and predict a present and future situation and therefore can be used for decision automation. They can subsequently perform an increasing number of tasks that are knowledge- and labour-intensive, ranging from search and translation to autonomously operating machines such as cars. This is a new situation that may lead to a deeper transformation through technologies than those seen during the first industrial revolutions, given that more intellectually demanding jobs could be affected. Still, change will be slow, for practical, legal and other reasons.

There is currently a major debate concerning the employment effects resulting from DDI. Many observers see a high risk that "smart" applications will further broaden the employment polarisation highlighted above, at least in the short run. More middle income jobs may be negatively affected – jobs largely held by the segment of the population that "glues" our societies together. Furthermore, DDI will also affect manufacturing by increasing labour productivity,[7] and that could reduce the number of blue collar jobs needed. So while manufacturing could return to OECD economies, at least to some extent, the extent to which this manufacturing on-shoring is likely to generate large numbers of jobs is equally debatable.

Furthermore, this trend may present emerging economies with new challenges, as their role as low-cost assembling points in global value chains may diminish. As a result, there is a risk that one of their historical routes to development – as well as their ability to leap-frog along it – could be reduced. Brynjolfsson and McAfee (2012a, 2014) and Cowen (2013) observe that after the last big recession in the aftermath of the global financial crisis that started in the year 2007, there was no detectable prospect of job recovery once economic growth took off again. More and more companies consequently announced they intended to replace jobs with machines. A prominent example is provided by the company Foxconn, one of the biggest companies for electronic products, which announced it would replace human workers by 10 000 robots in China (c|net, 2012; Stewart-Smith, 2012; Kan, 2013; Spiegel Online, 2014). Unfortunately, this development will not mean significantly more jobs in the developed world, either. And it has raised concerns among government in emerging economies including in China where Hon Hai Precision Industry is one of the largest private employers (Mozur and Luk, 2012).

The academic debate about technological unemployment will thus most likely intensify, especially since many of the technologies enabling DDI have still not seen large-scale deployment and their economic impacts are therefore still not fully known. In many ways, the world is today at the dawn of machine learning (ML – see Box 3.3 in Chapter 3 of this volume), at a development stage similar to that of the Internet in 1994: few practical commercial examples have reached maturity beyond their large-scale test phase, and much more that is now in the pipelines of research and development (R&D) labs is yet to come. But as applications develop quickly and become more cost-efficient, new generations of autonomous and semi-autonomous machines and systems will be deployed into every part of the economy, bringing with them the potential to displace work in these environments. This could theoretically lead to workerless factories, as the following sections suggest. Even if it causes only temporary friction in the economy, as Keynes once suggested, it is a development policy makers need to consider. Machine learning is as much about the competitiveness of the economy as it is about labour policy.

The following two sections illustrate how DDI-enabled applications will affect i) white collar and ii) blue collar jobs. As autonomous systems are increasingly able to perform intellectually demanding cognitive tasks including, as noted above, diagnosis of diseases based on the analysis of complex information and translation of complex documents and basic spoken language, questions emerge about the extent to which these systems can automate knowledge-intensive tasks that were until recently the object of more highly skilled white collar jobs. Furthermore, DDI enables a new generation of autonomous machines and robots with much more extensive capabilities that can be deployed and redeployed at much faster rates and more cost-effectively compared to current generation robots; these can affect manual labour-intensive blue collar jobs. Some have therefore argued that a long history of achievements of automated processes driven by data processing is reaching a threshold. The question however remains whether the effects on employment will lead to the replacement of jobs by machines, and/or to their "augmentation" or enhancement as better tools become available to the workforce (Davenport, 2014).

### *Data-driven decision automation affecting white collar jobs*

Data- and analytics-enabled smart applications are not used solely by Internet firms. The example of algorithmic trading systems (ATS) in finance, highlighted previously, now accounts for more than half of all financial market trades in the United States.[8] In health care, as another example, medical records with vital signs, magnetic resonance imaging (MRI) and other medical images, can now be analysed with each record representing a pattern that corresponds to diagnoses, therapies and treatments. Machine learning now makes it possible to develop (autonomous) artificial intelligence (AI) systems such as IBM's Watson that built on considerable parts of the worldwide knowledge base and are even capable of answering questions posed in natural language. Watson, which successfully competed on Jeopardy! against former winners, is now used to support medical diagnosis and therapy – building on more than 600 000 medical reports, 1.5 million patient records and clinical trials, and 2 million pages of medical journal text as of 2013 (IBM, 2013; Upbin, 2013). Such systems could have a profound impact on a number of jobs in the health sector, including high-skilled jobs such as radiologists and oncologists that involve spending a significant share (if not most) of work time identifying anomalies in medical images (Wang and Summers, 2012; Myers, 2011).

Similar systems have been suggested for other data- or information-intensive tasks such as legal advice and even legal decisions. Some of these systems are already used as the basis for legal analytical services provided by companies such as Lex Machina and Huron Legal. These systems can process millions of legal documents and retrieve the most relevant among them in seconds, and far more widely and thoroughly than people can do (Colvin, 2014). Even more striking is that some of these systems can actually predict court decision outcomes.[9] Katz, Bommarito and Blackman (2014), for instance, have developed a system that correctly identifies 70% of the United States Supreme Court's decisions and correctly forecasts 71% of the votes of individual justices across 7 700 cases, and more than 68 000 justice votes. It is difficult to imagine that these smart systems could replace lawyers and judges, although they may in some cases provide better advice "on whether to sue or settle or go to trial before any court and in any type of case" (Colvin, 2014). However, wide adoption of these systems could mean that young legal assistants may no longer be needed to search for relevant legal documents in the discovery phase of litigation, as many have been used to doing until today.

Even for data analysis, which is discussed below as a new key opportunity for employment, advanced analytic tools are now able to automatically fit thousands of statistical models to the available data and automatically generate and test different hypotheses (Davenport, 2014). A statistician relying on manual hypothesis testing can typically create only a few models per week. This has major implications not only for statisticians, but also for researchers (and their assistances). For example, King et al. (2004) presented a system that "automatically originates hypotheses to explain observations, devises experiments to test these hypotheses, physically runs the experiments using a laboratory robot, interprets the results to falsify hypotheses inconsistent with the data, and then repeats the cycle". Scientists are now further exploring the use of data analytics for automated hypothesis generation, and some have proposed analytical frameworks for standardising this scientific approach. Abedi et al. (2012), for example, have developed a hypothesis generation framework (HGF) to identify "crisp semantic associations among entities of interest". Conceptual biology, another example, has emerged as a complement to empirical biology; it is characterised by the use of text mining for automatic hypothesis discovery and testing. This involves "partially automated methods for finding evidence in the literature to support hypothetical relationships" (Bekhuis, 2006). Thanks to these types of methods, insights were possible which otherwise would have been difficult to discover. One example is the discovery of adverse effects to drugs (Gurulingappa et al., 2013; Davis et al. 2013).

### *New generation of autonomous machines affecting blue collar jobs*

Traditionally, robots have been used mostly in manufacturing where their speed, precision, dexterity and ability to work in hazardous conditions are valued. Traditional robots, however, were fast only in very precisely defined environments; setting up a robotic plant would take months if not years, to precisely plan all the movements of the robots down to the millimetre. Similarly, logistical robots that move the finished components have a precisely choreographed route. The robots might have sensors on board but most of the movements had to be pre-planned and programmed, which did not allow for much flexibility in the production of products. For this reason, the production of consumer electronics is still often done by hand, because the life cycle of consumer electronics and time to market is so short that the robotic factory would not be ready to make the current product by the time the successor should be on the market. This is radically changing because of DDI enabled by the IoT (Box 6.1), where sensor data are feeding machine-learning algorithms that often run via cloud computing services. As a result, machines are becoming more flexible and autonomous and can now perform a wider range of more complex manual work. To understand the employment implications, it is worth recalling the employment impact of the Jacquard loom, the first "programmable" mechanical weaving loom used on a large scale in the textile industry during the early 19th century (Box 6.2).

The potential of autonomous machines is best illustrated with autonomous vehicles such Google's driverless car, which collects data from all the sensors connected to the car (including video cameras and radar systems) and combines it with data from Google Maps and Google Street View (for data on landmarks and traffic signs and lights). If these autonomous vehicles are a success, then autonomous taxis, buses and trucks will be likely candidates for deployment. The effect could be that employment that in the past absorbed unskilled or low-skilled workers will no longer exist. There will still be jobs associated with providing these functions. However, many of them will require higher skills, for example for repairs and programming of robotic functions. Having a skilled labour force is therefore crucial.

Large warehouses have so far also been major employers of workers. In traditional warehouses, the workers walk with pick lists that indicate which items to pick. Modern warehouses use digital technology to direct workers to particular shelves and tells them what items to pick. The worker then scans the barcodes of the items picked and deposited. Workers walk many kilometres each day.[10] Other warehouses use conveyer belts for workers to put products on. The humans are controlled by the computer (see section below on "human computing", in particular Box 6.6). However, in some of the warehouses, the model of working has changed. In these warehouses the shelves are coming to the workers, carried by small driving robots such as those manufactured by Kiva Systems, a company acquired by Amazon after the latter started using Kiva's robots. It creates a different type of warehouse, where the workers stand still and the position of the shelves is dynamic. The location of the goods is continuously optimised, so that the most popular products are on the shelves that need to travel the shortest distance.[11] Pointing, a laser shows the worker what product needs to be picked and where it needs to be deposited. The effect is a supremely efficient warehouse that needs fewer workers to handle the same amount of orders.

### Box 6.2. **The Jacquard loom: A driver of industrial revolutions**

In 1801, Joseph Marie Jacquard, a French weaver and merchant, first demonstrated his more highly developed mechanical weaving loom, the Jacquard loom. Mechanical weaving looms had existed before, and had a breakthrough with the invention of the "wheeled shuttle" or "flying shuttle" by the English merchant and inventor John Kay in 1733 (Carlisle, 2004, Kessler, 2004). The key innovation of the Jacquard loom, however, was that it was controlled by an unlimited chain of replaceable punched cards, which enabled the Jacquard loom to be (re-)programmed for the manufacturing of a variety of textiles with different complex patterns. This was key to the Industrial Revolution of the late 18th century, in many respects:

1. The Jacquard loom can be seen as one of the earliest forms of software enabled technology. Punch cards were storage devices on which the woven pattern to be reproduced were encoded. Because the punch cards where replaceable, multiple patterns could be reproduced and if necessary combined to create even more complex patterns. Before the Jacquard loom, only plain (or at best extremely simple) woven patterns could be mass-produced by mechanical weaving looms. More complex patterns were only possible through manual labour.

2. The Jacquard loom had huge implications for higher-skilled textile workers as well. During the Industrial Revolution, the "traditional" mechanical weaving looms led to the automation of processes, rendering many jobs and skills in the "cottage industries" obsolete (*The Economist*, 2014; Dunne, 2014). Large quantities of textiles could be mass-produced at much faster rates than previously with manual workers, and with economies of scale that significantly reduced production costs. There was, however, one area where "traditional" mechanical weaving looms could not compete with skilled manual workers: the production of textiles containing extremely complex woven patterns and pictures. With the introduction of the Jacquard loom, however, even these tasks could then be performed by machines automatically, and at much the same rate and low costs as the production of textiles with plain woven patterns. As a result, even the more highly skilled textile workers, who once were not affected by automation, suddenly were no longer required in the production of complex woven patterns and pictures. As will be discussed below, a similar trend can also be expected to occur in the near future, with a number of middle income jobs being potentially affected by DDI.

3. Finally, the punch card inspired the first generation of computer storage devices, and is therefore considered an important step in the history of computers (Essinger, 2004), and an enabler of the digital (third industrial) revolution. The idea of punch cards inspired for instance the invention of the "mechanical tabulator" in 1889 (US Patent 395 782) by Herman Hollerith, an American statistician at the United States Census Bureau, who went on to become one of the founders of the company that later became known as the International Business Machines Corporation (IBM). Hollerith used this machine to encode data on punched cards more efficiently, thereby boosting data-processing capacities to a whole new level. Hollerith's tabulator enabled the US Census Bureau to complete its 1890 census within just one year, an operation that in the previous 1880 census had taken seven to eight years (Bruno, 2014). This meant a huge cost reduction for the bureau, that needed to employ more than 46 000 census clerks to collect the data – the cost reduction was estimated at the time to be around USD 5 million compared to manual tabulation (Aul, 1972).

C&S, a large supermarket wholesaler in the northeast of the United States, now operates a warehouse that is fully automated from the moment the pallets arrive from the manufacturer and the plastic wrap is removed, to the moment the pallet is put on the truck for transport to the supermarket.[12] Robots move through the shelves and pick the products, which then are handed to autonomously moving robots that bring the products to palletising robots. Not only is the robotic system faster and more efficient with less spillage, but it also allows for the building of higher and better pallets, decreasing the number of trucks needed to ship products, according to the system's manufacturer, Symbotic.[13]

The problems associated with building this kind of warehouse were mostly computational. Such problems include management of the movements of a few hundred robots, such that they do not have to wait for each other and can move at high speed, and that the breakdown of one robot does not break down the system. Another problem is the correct sequencing of picking products, so that a pallet contains products that are all near each other in the supermarket, but heavy products are at the bottom. Each pallet has to be calculated as a 3-dimensional object, not just by itself but also in relation to other pallets destined for the same truck and store. This is a computationally hard problem, known as the knapsack-problem. The solution to the problem came through work done on computer storage problems, where both in desktop computers and in cloud computing the optimal location for data has to be calculated.

In any case, the end result was a warehouse where one pallet assembly robot can palletise 600 cases per hour, compared to 150 for an experienced palletiser and 75 for a starting palletiser, and therefore an important productivity increase. Similar efforts have now also enabled robots to load trucks for package delivery. Today a robotic arm is capable of loading and even unloading the truck using sensors that measure where the other packages are and a three dimensional model that guides the loading of the truck.[14]

Workers are still not easily replaced. The picking of goods in Amazon's warehouse and the loading of trucks in the C&S warehouse continue to be done by humans.[15] But new robots are being introduced into the workplace that could change the situation. An example is the Baxter robot, a new robot that can work together with human workers. It can be programmed by the workers on the floor by moving its arms around, directing its tools and confirming each movement. The robot can then be programmed to perform a task such as packing or unpacking boxes, carrying items to or from a conveyor belt, counting them and inspecting them (Colvin, 2014). And it can easily be reprogrammed if the task needs to be optimised or changed. It is cheaper than comparable robots (USD 22 000) and can be programmed (again, on the job) in a matter of minutes, unlike traditional industrial robots that require days or weeks of highly specific programming by dedicated engineers. Tasks that the robot is currently performing include the boxing of goods and their transfer from one line to another. In such tasks, two robots can replace one human worker, though the robots do require oversight to the level of one supervisor per twenty robots. One could imagine that this robot will replace warehouse workers in jobs such as picking and packaging. Combined with robotic advances in manufacturing, the large deployment of these robots might one day lead to fully automated production processes, from design to delivery.[16]

## *Implications for skills and competencies*

The growing polarisation of employment described above has implications for the skills and competencies needed by the future working force. Studies that have looked in depth at the jobs that could be affected by the emerging automation opportunities also provide insights on the tasks that remain to be performed by humans and the skills and competence needed for these tasks. Furthermore, DDI provides opportunities for labour productivity growth not just through the reduction of jobs, but also through the "augmentation" or enhancement of existing jobs as better tools become available. In that respect, many of the tools described above will augment the intellectual and physical capacities of existing and new workers, opening a new range of possibilities for addressing current and future societal needs. The following sections discuss i) the mix of skills and competencies needed to perform the remaining and new labour-intensive tasks, and ii) the opportunities that DDI offers to augment and enhance the capacities of white and blue collar workers. Whether the future workforce will be adequately equipped with these skills and competencies will finally depend on the capacities of national education systems to support the development of these skills, as discussed further below in this chapter.

### *Skills and competencies needed for the remaining and emerging jobs*

The analysis of jobs affected by automation reveals that a number of tasks will remain to be performed only by humans, and that many of these tasks will become even more relevant for the workforce in the near future. Today, work that consists of following clearly specified directions is increasingly being carried out by computers and by workers in lower-wage countries where the labour costs are still below the costs of machines. The remaining jobs that pay enough to support families will require a deeper level of skills and competence.

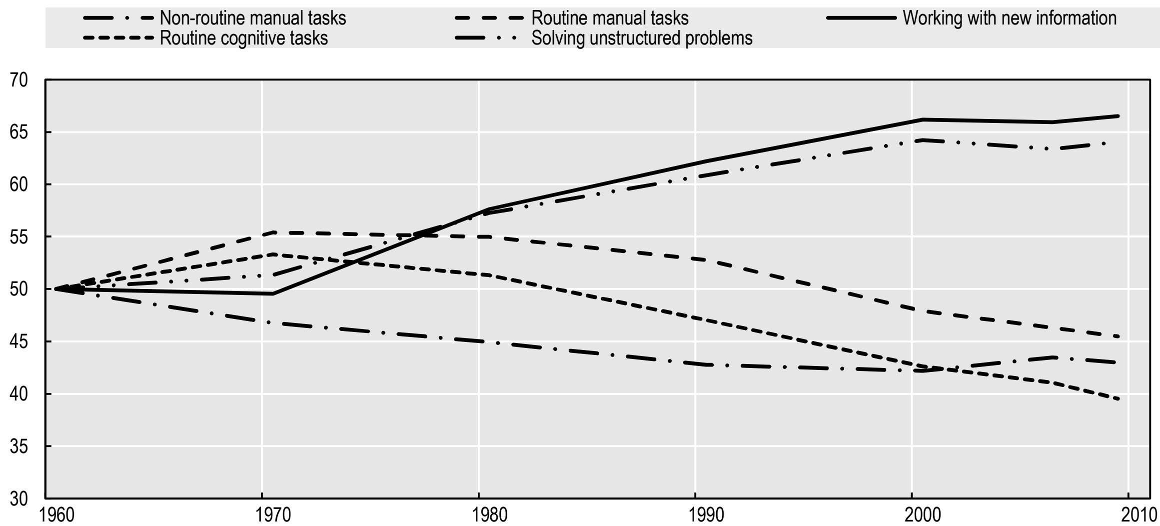
Based on the analysis by Autor and Price (2013), Levy and Murnane (2013) note three kinds of tasks on which labour markets will further concentrate (see Figure 6.3):[17]

- *Solving unstructured problems* – including tackling problems that lack rules-based solutions.
- *Working with new information* – including making sense of new data and information for the purpose of problem solving, decision making, or influencing the decisions of others. This includes many of the activities needed for DDI, as will be discussed in the next section.
- *Non-routine manual tasks* – Carrying out physical tasks that cannot be well described via rules because they require optical recognition and fine muscle control (advanced sensomotoric skills) that have still proved difficult for robots to perform.

While solving unstructured problems and working with new information will be particularly important for high-end jobs, carrying out non-routine manual tasks will become increasingly important for low-paying jobs (Levy and Murnane, 2013). This is in line with the observation of employment polarisation presented above – including the findings of Brynjolfsson and McAfee (2011, 2014), who state that a small group of people able to "race with machines" and a large group of people competing for lower-wage job opportunities could result from the DDI that we are witnessing at the moment.

Figure 6.3. **Index of changing work tasks in the United States**

Index, 1960 = 50




*Source:* Autor and Price, 2013.

Frey and Osborne (2013, pp. 24-27) provide a similar view on the areas of future employment left for humans, namely jobs that are less likely to be susceptible to computerisation. They argue with reference to other literature that three capabilities in particular remain difficult to automate. Based on the analysis of O*Net data for the United States, they estimate that around 47% of total employment in the United States does not rely on these capabilities, and is therefore seen by the authors as the theoretical maximum share of employment that could be negatively affected by automation. The three capabilities are:

- *Complex perception and manipulation* – Tasks that relate to an unstructured work environment.
- *Creative intelligence* – As creativity involves not only novelty but also values, creative thinking remains out of the realm of computers. Furthermore, creative intelligence is often associated with human intuition as a genuine human capability.
- *Social intelligence* – includes among others the real-time capacity to recognise human emotions and the ability to respond intelligently to such inputs. For computers, this remains a challenging problem.

Elliott (2014) offers a detailed examination of the occupations that are more likely to be affected by automation, based on analysis of the clusters of skills required by these occupations. These clusters include:

- *"Vision movement"* – This includes the combined capacity of i) recognising objects and different features of those objects, including their position in space (vision), and ii) spatial orientation, co-ordination, movement control and body equilibrium. Physical movement is important for jobs in construction, maintenance and production, as well as for jobs in food and personal service. These two large occupational groups represent 30% of current employment.

- *"Language reasoning"* – Including the capacity to deal with natural language: understanding speech, speaking, reading and writing, combined with the capacity to reason, including recognising that a problem exists, applying general rules to solve a problem, and developing new rules or conclusions.

Elliott (2014) suggests that occupations less affected by automation are those that require a high level of at least one of the clusters of skills highlighted above, although the author acknowledges that the progress in automation will soon make even these occupations susceptible to automation in the middle to long term. Elliott (2014) therefore concludes that occupations that involve higher levels of language and reasoning skills are *currently* beyond the capabilities of automation. These include occupations related to education, health care, science, engineering and law.[18] However, as highlighted above, machine learning now makes it possible to develop (autonomous) artificial intelligence (AI) systems such as IBM's Watson that are capable challenging humans even for health care and legal jobs involving higher levels of language and reasoning skills.

### *Augmenting humans' capacities – Dancing with the machines*

As highlighted above, employment opportunities will remain for people with the right mix of skills and competencies. DDI could help augment the intellectual and physical capacities of individuals for these opportunities. Using ever more powerful technical systems enabled by data and analytics as input into the contribution of human work, DDI can further enhance human creativity, social intelligence and sensomotoric skills – and thus combine the experience-based capabilities of analytics with the cognitive capabilities of a highly educated human. This potential has been highlighted by several authors. Brynjolfsson and McAfee (2014), for instance, suggest that we have to learn to race *with* the machines, instead of against the machines, by adding intuition and creativity to the capabilities of new developments driven by big data. Cowen (2013) predicts that the highest performance will be achieved by "freestyle teams", where humans take advantage of their specific know-how and their intuition to best use and connect several systems to get the best results.[19] Levy and Murnane (2013) call this "dancing with robots".

The example of chess is often given to illustrate the power of "dancing with the machines". Currently, even chess software implemented on a smartphone is strong enough to beat most human chess players. In 1997, Deep Blue, a chess-playing computer developed by IBM, defeated the world champion Garry Kasparov. But when it comes to so-called freestyle chess competitions, it is neither humans nor computer systems that win, but the combination of computer systems working with a team of humans. Experience also shows that those humans in "freestyle teams" do not have to be high-level chess players themselves. Their specific know-how is about weaknesses and excellence of all the specific computer systems and how to work with them in a fast and flexible way. They use outputs of systems as an input to other systems, varying and filtering the results. In that way, a network of computer systems is used to derive a viable proposal for the next move to be made in the running chess play. All these computerised decisions are based on data and analytics used by humans with a sufficient level of skills in data and analytics (i.e. data specialist skills – see next section).

The use of data analytics with machines can complement humans' creativity in finding new ways forward and solutions to problems. But as highlighted by Brynjolfsson and McAfee (2014), the combination of humans with machines is also important because at the end of the day it will be humans that will have to take the responsibility for decisions made, as no machine can be held accountable for false decisions. This is important also because as Taleb (2005, 2010) has demonstrated, “black swans” could lead to false automated outcomes. In other words, many data-driven decisions are based on statistical learning and assume that statistical distribution patterns can always adequately model the reality, including in the future (see Chapter 3 of this volume).

However, many examples show that this assumption does not always hold true; severe economic and social consequences are sometimes the result. Financial crises and the complete failure of modern economics to anticipate the outbreak of the global crisis in 2007-08 can be seen as one example. In medicine, there are cases where the medical data and images used as the basis for diagnostics are not enough to make a correct diagnosis. Professional high-skilled physicians with experience, creativity and intuition and personal interaction with the patient are therefore needed to make the best decisions. Data analytics-empowered machines would provide helpful tools to support their decisions. Those humans teaming with the machines and taking over the accountability and responsibility for their decisions must be highly educated, however. This is because there may be situations where the machine contradicts the opinion of the human decision maker, raising the question whether humans are willing and able to take over the responsibility when overriding a machine’s suggested decision (see Box 6.3). Instead of leading us into a future of human-machine collaboration, the future could be a “domination of empiricism” or a “dictatorship of data”, where less educated or less concerned decision makers automatically follow the decisions of machines (Mayer-Schönberger and Cukier, 2013).

Box 6.3. **What is new for decision makers with big data?**

In order to understand the implications of big data for decision making, it is important to distinguish big data from conventional information processing. Two major differences deserve highlighting here:

1. Big data is often about providing a kind of “manifest what” by extracting value from a flat and unstructured “datafied universe of information-shreds” with unknown veracity. This helps answer questions on the basis of “alleged insight” via calculated approximations and correlations.
2. Conventional data analytics is about providing a kind of “know what” instead of “manifest what” by extracting “know why” as value. This helps answer questions on the basis of “explicit insight” via revealed causation.

The difference is thus mainly that between correlation and quantitative reasoning compared to causation and qualitative reasoning. The observed trend “from causation to correlation” deserves policy makers’ attention where it is leading to “liability aversion”. In some cases liability aversion can lead to huge social costs – for instance, when financial risk assessments are based solely on correlations.

Source: Herlyn et al., 2015.

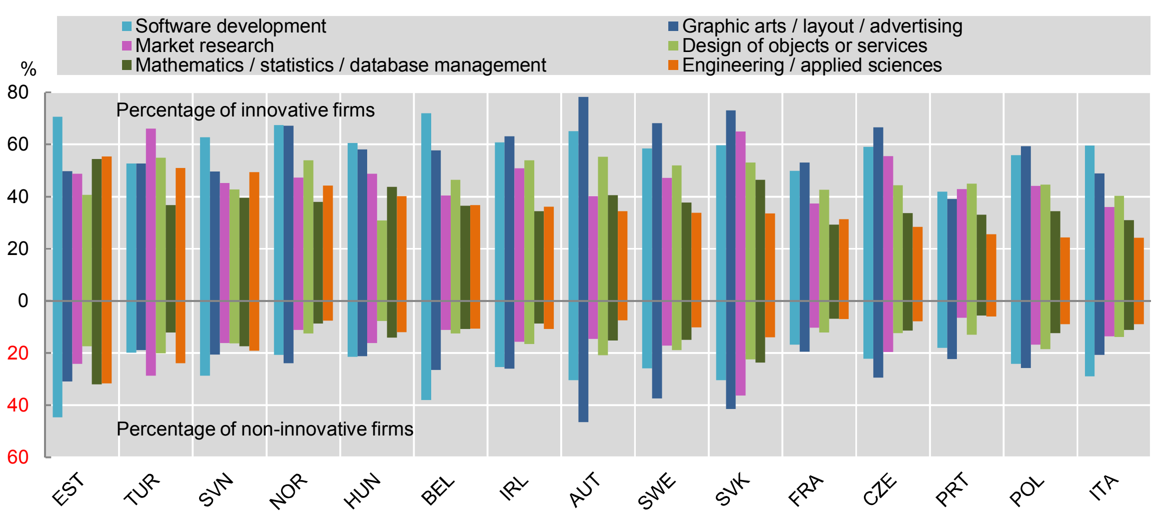
## 6.2. The growing importance of data specialist skills and employment

As highlighted in Chapter 2 of this volume, DDI is furthering the creation of new businesses and business models, many of which were unheard of a decade ago, such as data analytic service providers and data brokers and explorers. The results are employment opportunities across the economy. Along with the interest in extracting insights from huge collections of data, the need particularly for people who inherit the skills required for extracting insights from data is growing. This is confirmed by Levy and Murnane (2013), according to whom "working with new information", including making sense of new data, is one of remaining growing task categories for which labour demand can be expected to increase in the future (see Figure 6.3). Data specialist skills are thus critical for the workforce to be able to "dance with the machines", as discussed above. Furthermore, the right skills in data analytics are essential to understand how to appropriately use data and analytics, and how to deal with the limitations of data-driven decision making highlighted in Chapter 3.

In that respect, data specialist skills are also a key enabler of DDI, as confirmed by business innovation surveys showing that firms using (internal or external) skills related to data and analytics (i.e. mathematics, statistics and database management skills) are more likely to innovate (Figure 6.4).[20] Evidence suggests furthermore that firms with better access to data specialist skills are more likely to gain faster productivity growth through DDI. A recent study by Tambe (2014) was based on an analysis of 175 million LinkedIn user profiles, out of which employees with skills on big data-specific technologies were identified. The study indicates that firms' investment in big data-specific technologies were associated with 3% faster productivity growth, but only for firms that i) already had access to significant data sets and ii) were well connected to labour networks with sufficient expertise in big data-specific technologies. (The estimated output elasticity of 3% resulted after controlling for firms' adoption of data-driven decision making.) This highlights the complementarity effects among data, analytics and skills, the understanding of which merits further study.

Figure 6.4. **Firms using innovation-relevant skills, 2008-10**

As a percentage of innovative and non-innovative firms




*Source:* OECD Science, Technology and Industry Scoreboard 2013, http://dx.doi.org/10.1787/888932890770.

However, some evidence suggests that the demand for data specialist skills already exceeds the supply on the labour market. The Economist Intelligence Unit (2012) survey shows that "shortage of skilled people to analyse the data properly" is indicated as the second biggest impediment to making use of data analytics. For consumer goods and retail firms it is the single biggest barrier, cited by two-thirds of respondents from those sectors. Some other studies have concluded that there are considerable mismatches between the supply of and demand for data specialist skills. MGI (2011), for instance, estimates that the demand for deep analytical positions in the United States could exceed supply by 140 000 to 190 000 positions by 2018. This does not include the need for an additional 1.5 million managers and analysts who can use big data knowledgeably. However, further evidence to confirm this trend across countries is needed.

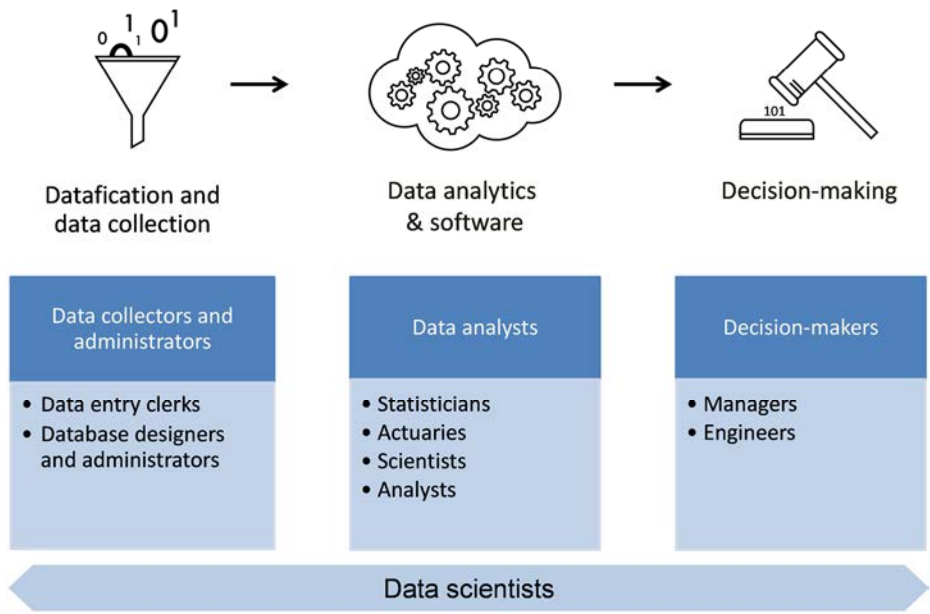
After defining data specialist skills, this section looks at the availability of these skills across the economy and reviews trends in their development over time, including wages and vacancies. Where possible, the section compares these developments with broader economy-wide trends in employment, focusing on ICT specialist employment. Following the methodology applied in OECD, 2013a, insights are provided on the data intensity and diffusion of DDI across the economy from the perspective of skills and employment, and offers a picture of the relative demand for data specialist skills across sectors, highlighting those that intensively employ data specialists. By measuring the share of each sector's workforce related to data, the section provides an approximate figure for the use of data (and analytics) across the economy, a figure that until now could not be provided through official statistics for reasons discussed in Box 2.1 in Chapter 2 of this volume.

### *Defining data specialist skills and employment*

There is currently no commonly adopted definition of data specialist skills. To a large extent this is due to the fact that these skills have not received much attention in literature compared for example to ICT (specialist) skills – which in many respect cover, although not perfectly, what are termed in this chapter data specialist skills. Furthermore, DDI is a relatively new phenomenon that has only recently caught the attention of decision and policy makers, and it is therefore only recently that data specialist skills have risen to the top of the agenda of different stakeholders. And, last but not least, DDI is not only a new phenomenon but also a rapidly evolving one, which means that many data specialist jobs are rapidly evolving as well. Whereas some of these jobs are already established in labour markets, many – such as "data scientists" – are rather new professions that just recently emerged in light of a convergence of disciplines, including computer science and statistics but also natural and social sciences as well as business management, marketing, and finance.

All these data specialist professions have one thing in common: *working with data constitutes a main part of their job*. Different data specialist occupations can be identified using the data value cycle introduced in Chapter 1 as a framework (Figure 6.5). These include in particular occupations that mainly i) collect and/or manage data, such as data entry clerks and database designers and administrators, and ii) analyse data through analytics, including in particular statisticians and actuaries; scientists such as astronomers, epidemiologists and economists; and analysts such as those in finance, market research and intelligence. But it also includes related associate professionals and clerks such as statistical assistants. To a limited extent, data specialist professions also include iii) data-driven decision makers such as managers and engineers, for whom however working with data rarely constitutes a major portion of their jobs.

Figure 6.5. **Main phases of the data value cycle with their key types of data specialist occupations**




Given this definition of data specialists, measuring related occupations based on official statistics remains challenging for several reasons. First, the rapidly evolving nature of data specialist skills has led to the emergence of new professions such as "data scientists" that are not properly captured by official statistics (Royster, 2013). Also, some occupations such as economists may often, but not always, require working with data as a main part of the job. And finally, the official national statistics provided are often not granular enough to really capture data specialist occupations, and this becomes even more of an issue when comparing available statistics across countries. Box 6.4 therefore proposes an operational cross-country definition of data specialist skills that, for comparability and measurement reasons, excludes a number of occupations that would otherwise be captured by the framework presented in Figure 6.5.

Box 6.4. **Data specialists: Towards an operational cross-country definition**

Following the OECD definition of ICT specialist, data specialists are defined for the purpose of this report as those occupations for which *working with data constitutes a main part of the job*. In an attempt to provide comparable measures across OECD countries, data specialists have been defined according to the 2008 International Standard Classification of Occupations (ISCO-08) to include the following two occupations at three-digit level:

- 212 – Mathematicians, actuaries and statisticians
- 252 – Database and network professionals.

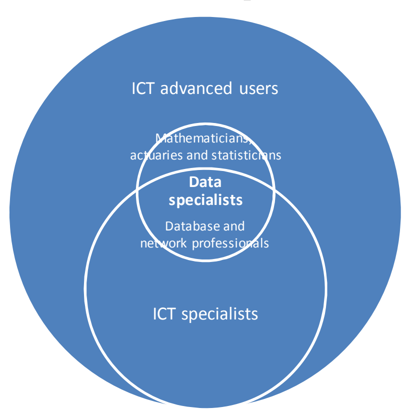
### Box 6.4. **Data specialists: Towards an operational cross-country definition (cont.)**

It should be noted that some occupations considered as data specialist are missing. These include in particular "data entry clerks" (4132), which is only available within a larger set of occupations (at 3-digit level) that additionally includes non-data specialists.[1] The same is true for "statistical, mathematical and related associate professionals" (3314), which are also a subset of a much larger non-data-specialist group of occupations.[2] Furthermore, "physicists and astronomers" (2111) are often considered data-intensive occupations, but the 3-digit occupation "physical and earth science professionals" to which they belong not only includes meteorologists (2112) which can be considered data-intensive, but also chemists (2113). Epidemiologists are not explicitly captured by ISCO-08, but are part of a larger group including "biologists, botanists, zoologists and related professionals" (2131). Finally analysts often also include a large number of non-data specialist occupations such as "advertising and marketing professionals" and "public relations professionals", some of which rarely require working with data as a main part of the job.[3]

The proposed definition, including 212 and 252, is therefore considered to best strike the balance in capturing the employment activities related to the use of data across countries, and is seen as a *narrow definition*. With the help of correspondence tables, a comparable list of occupations has been compiled using regional classifications to measure data specialist employment in Australia, Canada and the United States (see Annex Tables 6.A3, and 6.A4). A much *broader definition*, including those occupations highlighted above, deserves further study.

Finally, it is interesting to note that the definition of data specialist proposed here is not a perfect subset of the OECD definition of ICT specialists (which includes database and network professionals), but the definition of data specialist further includes "advanced ICT users" (mathematicians, statisticians and related professionals) that were part of a broader definition of ICT skill occupations presented in OECD, 2005 (see Figure 6.6)

Figure 6.6. **Data and ICT specialists in context**




1. These include "typists and word processing operators" (4131). It should be noted that the omission of these occupations from an operational cross-country definition should not provide any significant challenge, given their declining share in the economy.

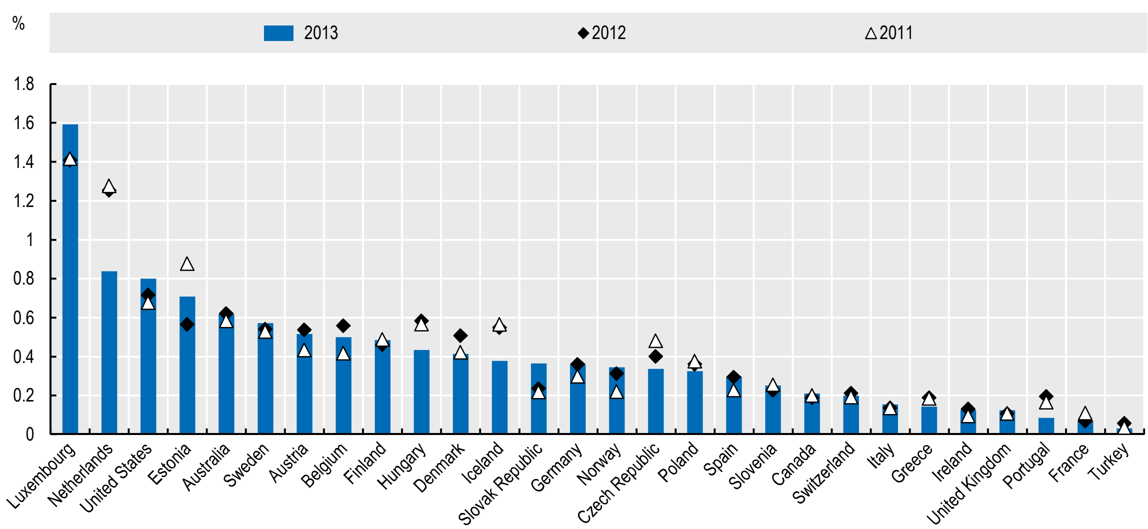
2. The group includes "securities and finance dealers and brokers" (3311), "credit and loans officers" (3312), "accounting associate professionals" (3313), and "valuers and loss assessors" (3314).

3. They include "accountants, financial and investment advisers" and "financial analysts" (241), which can be considered more data-intensive, but also "technical and medical sales professionals" and "information and communications technology sales professionals" (243), which are rarely considered data-intensive.

Estimates based on the definition proposed in Box 6.4 suggest that data specialists in 2013 accounted for over 0.6% of total employment in countries such as the Netherlands, the United States, Australia and Estonia, while in Luxembourg the share of data specialists almost reached 1.6% of total employment (Figure 6.7). In countries such as Portugal, France and Turkey, the share of data specialists is far below 0.1%. In most economies, the share has increased significantly over the past years, suggesting not only that demand for data specialists has increased faster than demand for other types of jobs, but also that these economies have become more data-intensive over time. Employment figures for Canada and the United States show that the share of data specialists in total employment has rapidly increased since 1999, even faster than the share of ICT specialists (Figure 6.8; see Figure 6.9 for Canada). For comparison, ICT specialists in the 28 OECD countries for which data are available corresponded to almost 3.5% of total employment in 2013 (about 14.1 million jobs). There are however some notable exceptions to the growing share of data specialists, namely in the Netherlands, Hungary, Iceland, Denmark, Czech Republic, Poland, Greece, and Portugal.[21]

Figure 6.7. **Data specialists in selected OECD countries, 2011-13**

As share of total employment

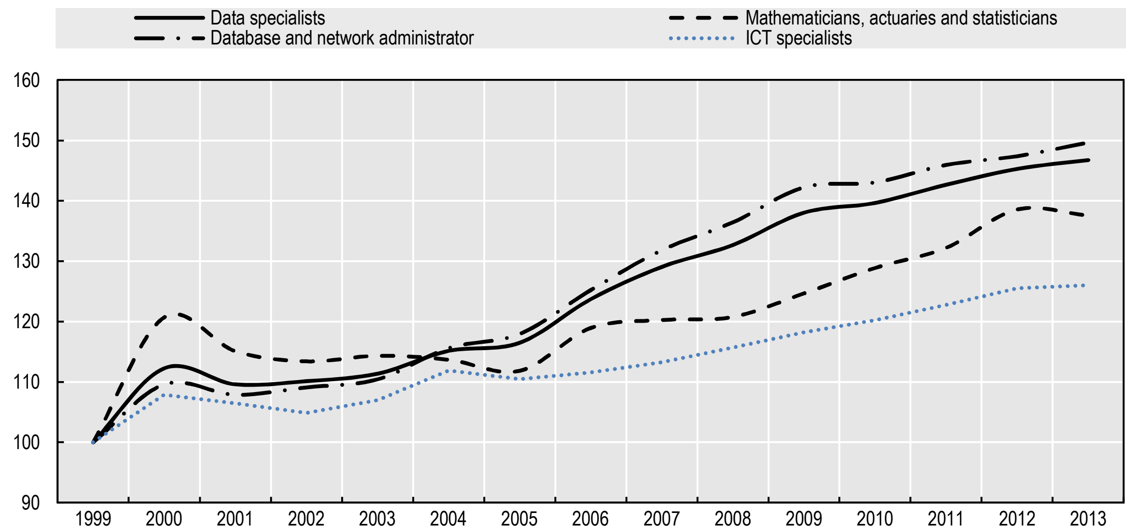



*Note:* Data for Ireland and the United Kingdom only include ISCO-08 code 212 “mathematicians, actuaries and statisticians” as data for code 252, “database and network professionals”, are not available. Data for Canada include the equivalent of ISCO-08 codes: 212 and 252. Data for the United States are overestimated since parts of other ISCO-08 codes (3514 and 2519) are included.

*Source:* Based on data from Eurostat, Statistics Canada, Australian Bureau of Statistics Labour Force Surveys and US Current Population Survey, March Supplement, February 2015.

Figure 6.8. **Trends in the share of data specialists in the United States, 1999-2013**

Index, 100 = 1999, share in total employment

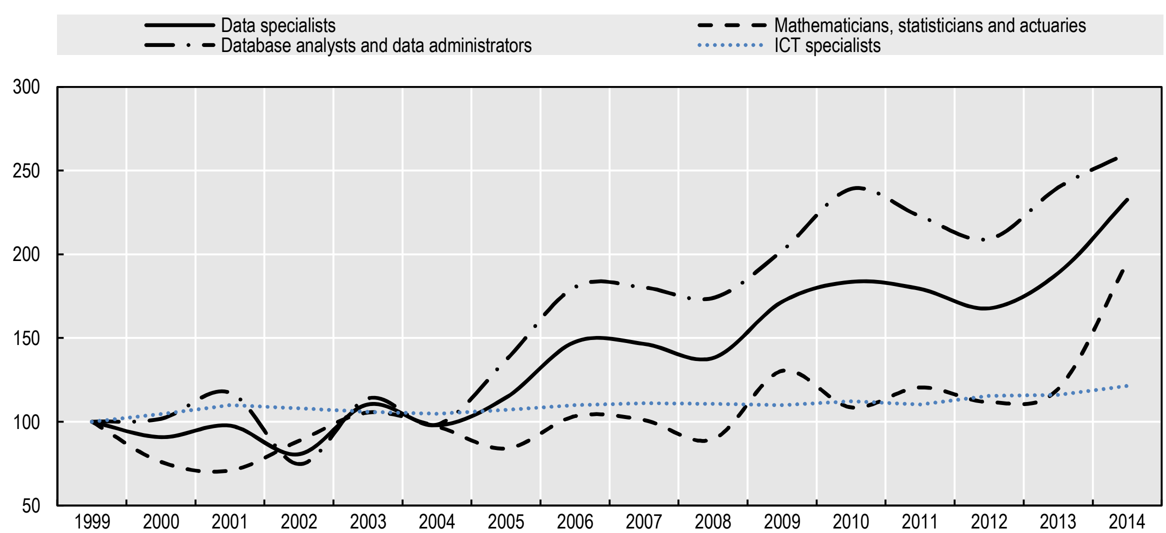



*Note:* "Data specialists" does not correspond here to the ISCO definition presented in Box 6.4. In order to be consistent across years, the definition has been slightly modified and does not include "information security analysts" (SOC 2010 code 15-1122), "computer network architects" (15-1143) or "computer occupations, nec" (15-1199).

*Source:* Bureau of Labor Statistics, Occupational Employment Statistics (OES), www.bls.gov/oes/home.htm, November 2014.

Figure 6.9. **Trends in the share of data specialists in total employment in Canada, 1999-2014**

Index, 100 = 1999

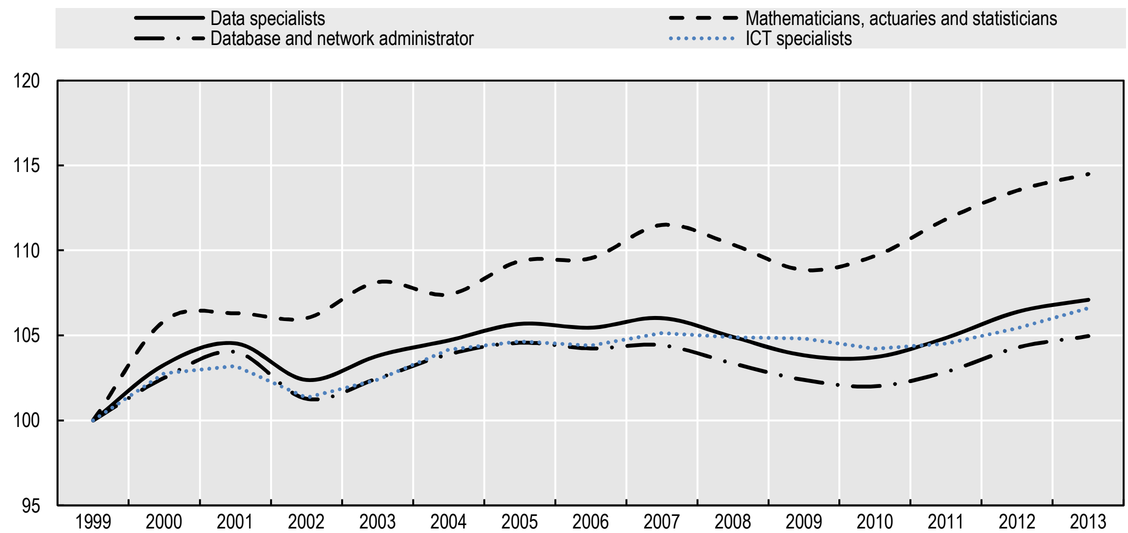



*Note:* "Data specialists" does not correspond here to the ISCO definition presented in Box 6.4. In order to be consistent across years, the definition has been slightly modified, and only includes ISCO 08 code 212, "mathematicians, actuaries and statisticians", and code 2521, "database designers and administrators" (equivalent to NOCS 2011 code 2172, "database analysts and data administrators").

*Source:* Statistics Canada, labour force survey, February 2015.

Figure 6.10. **Trends in relative average wage of data specialists in the United States, 1999-2013**

Index, 100=1999

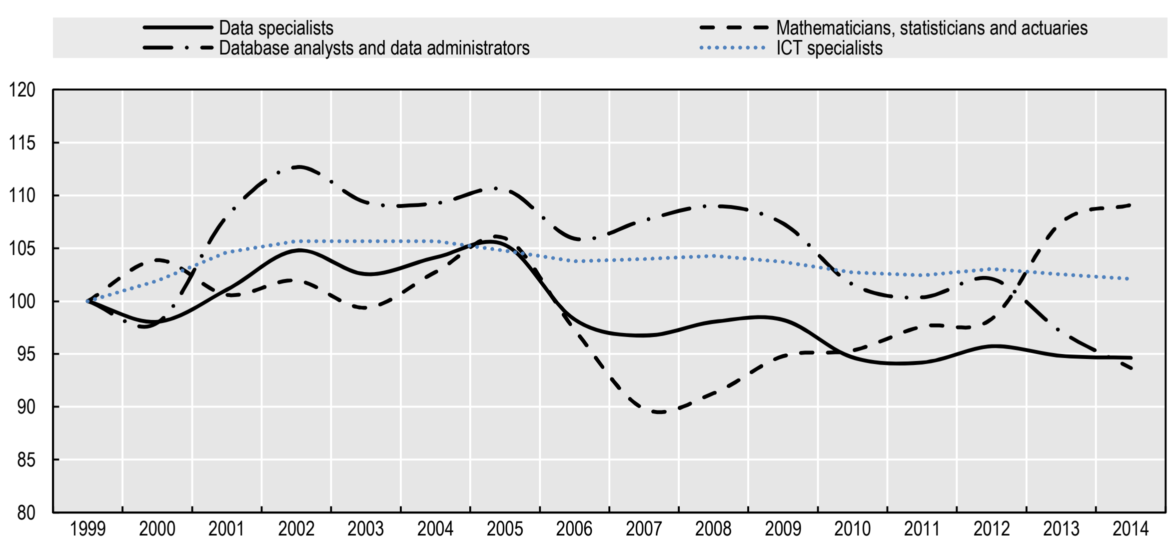



*Note:* "Data specialists" does not correspond here to the ISCO definition presented in Box 6.4. In order to be consistent across years, the definition has been slightly modified and does not include "information security analysts" (SOC 2010 code 15-1122), "computer network architects" (15-1143) or "computer occupations, nec" (15-1199).

*Source:* Bureau of Labor Statistics, Occupational Employment Statistics (OES), www.bls.gov/oes/home.htm, November 2014.

Figure 6.11. **Trends in relative average wage of data specialists in Canada, 1998/99-2013/14**

Index, 100 = 1998/99



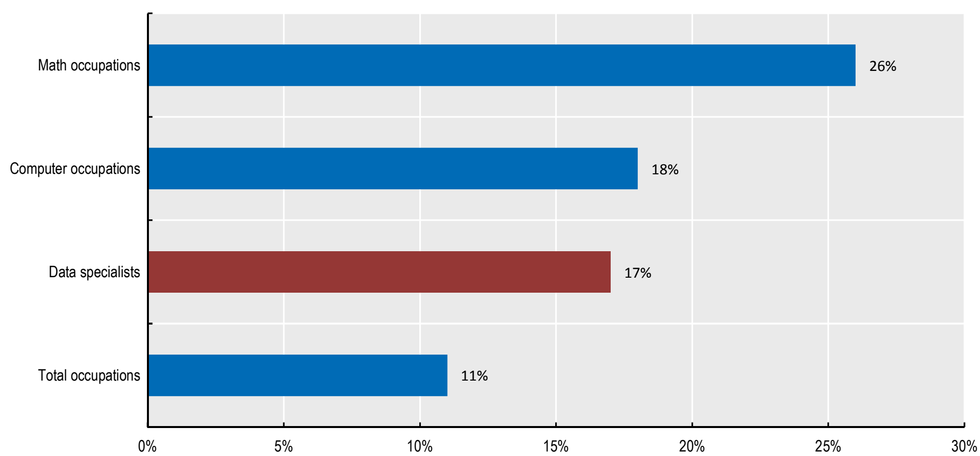

*Note:* Wage data are a two-year moving average based on average weekly earnings from 1998/99 to 2013/2014. "Data specialists" does not correspond here to the ISCO definition presented in Box 6.4. In order to be consistent across years, the definition has been slightly modified, and only includes ISCO 08 code 212, "mathematicians, actuaries and statisticians", and code 2521, "database designers and administrators" (equivalent to NOCS 2011 code 2172, "database analysts and data administrators").

*Source:* Statistics Canada, labour force survey, February 2015.

It is worth noting that the ratio of average annual wages of data specialists to that of all occupations has remained relatively stable in the United States in the last decade (Figure 6.10, for Canada see Figure 6.11), suggesting that demand for data specialists was satisfied through labour markets in general. Based on estimates of the US Bureau of Labor Statistics, demand for data specialist jobs are however expected to grow to 17% in the United States between 2012 and 2022 (Figure 6.12). This is six percentage points faster than the estimated total employment growth for that same period. Statisticians, actuaries and mathematicians are expected to have the fastest growth between 2012 and 2022 (26%). This is consistent with the observation that statisticians, actuaries and mathematicians are expected to be in higher demand as DDI becomes more important for businesses. These occupations have also seen the fastest growth in relative wages since 1999, compared to data specialists and ICT specialist for which relative wages have grown more modestly (Figure 6.10). But it is also observable that the share of statisticians, actuaries and mathematicians has been decreasing since 2012, suggesting – along with the further growing relative wages for that group – that the United States could be facing a shortage.

Figure 6.12. **Data specialist jobs outlook in the United States, 2012-22**

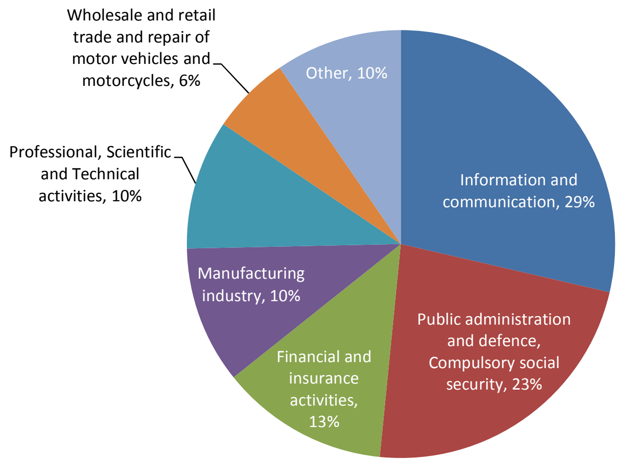
Percentage change in employment, projected 2012-22




*Source:* Based on the US Bureau of Labor Statistics, Employment Projections programme, December 2014.

Data specialists are more likely to be in highest demand in those economies where data-intensive industries are more prevalent, such as in Luxembourg where the financial sector is a major industry. The most data-intensive industries employing the highest share of data specialists are still the ICT service industries (see Figure 6.13),[22] and in particular i) IT and other information service industries, but also ii) insurance and finance, iii) science and research and development, iv) advertising and market research, as well as v) the public sector including extraterritorial organisations and bodies (such as the OECD, UN and other international organisations). A similar concentration can be observed for the United States. These findings are in line with recent studies suggesting that ICT firms are still leading in the use of advanced data analytics; according to Tambe (2014), only 30% of Hadoop investments come from non-ICT sectors, including in particular finance, transportation, utilities, retail, health care, pharmaceuticals and biotechnology. It is interesting to note that most of the data-intensive sectors also tend to have a high ICT intensity (ICT expenditure as a share of output).[23]

Figure 6.13. **Distribution of data specialists per industry in selected OECD countries, 2013**




*Note:* Industries are based on ISIC rev. 4 (2-digit codes). OECD 25 includes all OECD EU Member Countries plus Iceland, Norway, Switzerland and Turkey.

*Source:* European Union Labour Force Survey, November 2014.

## *Data specialists in transformation*

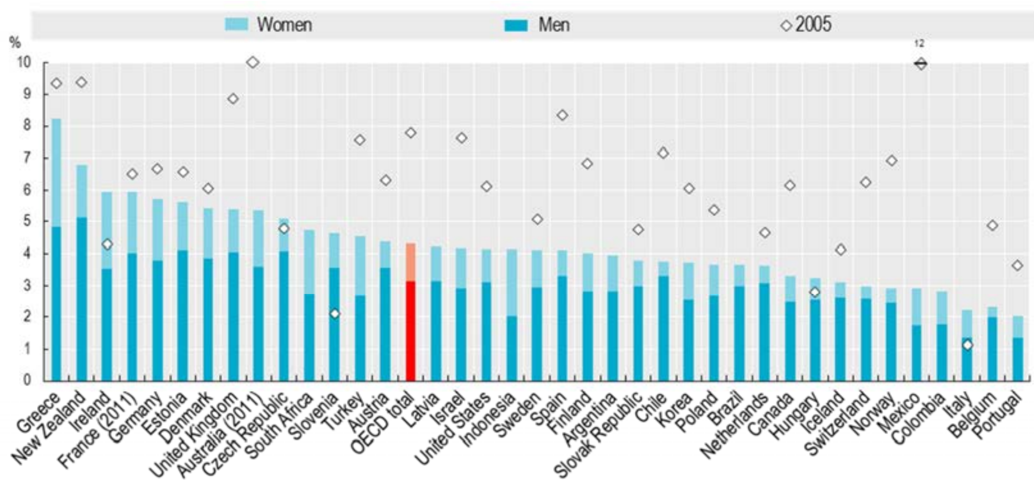
There are a number of sociological and technological trends that are transforming data specialist occupations. One important sociological trend is the increasing participation of women in data specialist occupations, which has not only contributed to a higher diversity, but also led to a higher supply of skills in the labour market (see Box 6.5). Another important (technological) trend is the increasing role of ICTs in data specialist occupations, and in particular the impact of the convergence of disciplines such as mathematics and statistics, but also that of natural and social sciences – such as physics and economics – with computer science. This convergence has led to a new class of data specialist, the "data scientist", a term that is still vague but used by several authors to describe the emergence of a "new" discipline, job category or career path that has grown in importance and prominence with big data. The major data specialist occupations are discussed in more detail in the following sections, beginning with traditional occupations that include data entry clerks and database designers and administrators, and then actuaries, mathematicians and statisticians. The section ends with a discussion of the term "data scientist" itself.

## Box 6.5. **The growing importance of women in data specialist occupations**

Women still participate significantly less in the data specialist occupations than men. One out of five data specialists is a woman. However, while their share in overall employment remained stable (0.05%) between 2011 and 2013, the proportion of women increased from 16% in 2011 to 20% in 2013. When looking at the occupations included in the definition of data specialists, it is interesting to note that around 40% of "mathematicians, actuaries and statisticians" are women.[1] The picture is somewhat different for the "database and network professionals", where only 14% are women.

The still relatively low share of women in data specialist occupations is reflected in the low participation of women in tertiary-level graduate programmes related to mathematics, statistics and computer science. In 2012, for instance, the share of data-related graduates among all tertiary graduates in the OECD area was 4.3%, of which 27% were women (Figure 6.14).

Figure 6.14. **Data-related tertiary graduates, by gender, 2005 and 2012**
As a percentage of all tertiary graduates




*Note:* "Data-related tertiary graduates" has been defined for this figure as persons who have attained a degree in the field of Computer Science (ISC 48) and Mathematics and Statistics (ISC 46) based on the International Classification of Education (ISCED-97), levels 5A and 5B. Data for Luxembourg and Japan are not available, nor are data for Australia or Israel with respect to ISCED 5B data.

*Source:* OECD Education Database and OECD (2014f), *Education at a Glance 2014: OECD Indicators*, OECD Publishing, Paris, www.oecd.org/edu/eag.htm, accessed 28 June 2015.

Evidence suggests that female data specialists tend to concentrate in sectors such as "public administration and defence"; "compulsory social security" represents almost 20% of all female data specialists, followed by "financial and insurance activities" (17%). "Computer programming, consultancy and related activities" and "information services activities" industries also attract a good number of women (16%). By comparison, male data specialists are concentrated in "computer programming, consultancy and related activities" (over 23%), followed by "financial and insurance activities" (12%). "Public administration activities" comes third, with near 10% of all male data scientists.

1. In 2014, half of the "mathematicians, actuaries and statisticians" were women in Australia. In Canada, the percentage is 44% for the same year.

### *Traditional data specialist occupations*

The main responsibilities of data specialists are the collection, management, processing and analysis of data sets to discover new insights – for example, through statistical modelling – to enable better decisions and predictions about the future. Traditionally these functions have often been undertaken through a division of labour among different occupations along the data value cycle presented in Figure 6.5.

*Data entry clerks* often stand at the beginning of the data value cycle, given that their main tasks are to collect and enter data into "electronic equipment, computerised databases, spreadsheets or other data repositories using a keyboard, mouse, or optical scanner, speech recognition software or other data entry tools" (ILO, 2009). Data entry clerks are especially needed to transform and store unstructured data (i.e. data that have no predefined data model or structure, such as text and multimedia files) in structured digital data formats. To understand the historical importance of data entry clerks, it is necessary to recall that unstructured data are estimated to still account for between 50% and as much as 85% of all data stored in organisations (see Chapter 3 of this volume).[24] And many data-intensive operations, such as a census, were and still are not possible without data entry clerks (see the discussion of Hollerith's tabulator in Box 6.2).

However, with rising computing capacities, data analytics is increasingly able to automatically extract some structures embedded in unstructured data, including multimedia content (Chapter 3).[25] As a result, data entry clerks have become less and less important for most organisations, a trend reflected in the ever decreasing share of data entry clerks observed in the United States since 1999. In addition, the increasing deployment of sensors with the Internet of Things has significantly increased the potential to automate data collection and storage to such an extent that few domains will remain where data entry clerks will be needed in the near future, including for the collection of census data (see Reimsbach-Kounatze, 2015). In that respect, data entry clerks can be seen as one of the job categories that DDI has successfully rendered less important; ironically, these jobs were initially one of the key enablers of DDI. As will be highlighted further below, this ironic twist is not limited to data entry clerks but also includes more intellectually demanding data specialist jobs, such as even statisticians.

Although the decreasing share of data entry clerks can be seen as an indicator for the increasing capacity of machines to do their jobs, their full replacement by machines is still some way off. As highlighted in the previous section, the capacity of "language reasoning" (Elliott, 2014) remains the competitive advantage of humans over machines, and this capacity has been one of the key reasons why data entry clerks remain employed today. Language reasoning includes the capacity to recognise meanings, which is essential – for example, when extracting related but ambiguous information where different ways of representing the same entity prevents computers from making semantic linkages. The problem is particularly relevant when different data sets need to be linked. The demand for these types of data-related tasks is reflected in the increasing number of crowdsourcing activities that some have referred to as "human computing", and that are offered through services such as Amazon Mechanical Turk (MTurk) since 2005 (see Box 6.6). These services involve small tasks for which human intelligence is required and no cost-efficient algorithm exists; examples include data cleaning and verification, including the classification of data entities and the identification of duplicates.

### Box 6.6. **Crowdsourcing of human intelligence tasks: "Human computing" and "micro tasking"**

While computing and automation technologies are steadily improving, there are still many tasks that human beings can do much more effectively than computers, such as identifying objects in a photo or video, performing data de-duplication or transcribing audio recording. To perform these often one-time tasks, firms tend to hire temporary workers. Crowdsourcing, a workforce for human intelligence tasks (HITs), is increasingly used as alternative to solve this problem while providing firms with even more flexibility and scalability when outsourcing these labour-intensive tasks that computers cannot perform. This process is often referred to as "human computing" to illustrate the reverse role between humans and computers, where computers "use" humans to solve problems that computers cannot perform. But the term also refers to the more traditional term "computer", referring to a human that "is supposed to be following fixed rules; he has no authority to deviate from them in any detail" (Turing, 1950).

Amazon is still the most prominent large-scale provider of "human computing" services over the Internet, since it launched its crowdsourcing marketplace for digital work called Amazon Mechanical Turk (MTurk) in 2005. Requesters advertise small projects that cannot be fully carried out by computers on the online platform. Worker called "turkers" can then complete those one-time tasks for mostly a very small amount of money, usually ranging from as little as USD 0.01 for a quick task up to rarely more than USD 100 for more complex jobs. Currently, there are around 500 000 workers from 190 countries registered at Amazon MTurk. Especially for people living in developing countries, MTurk and similar services have been highlighted as a job opportunity to overcome poverty, although the requirements of an Internet connection and English language skills still restrict this potential. Samasource, a nonprofit organisation based in the United States whose mission is "to use work, not aid, for economic development", provides data-related services to large companies in the United States and Europe. It divides the work up into small pieces (called "microwork") and then sends it for completion to delivery centres in developing regions including Haiti, India, Kenya and Uganda (Gino and Staats, 2012).

While they represent job opportunities for some, MTurk and similar services such as Samasource have been criticised, analogous to the socially unacceptable working conditions in the textile industry in the 19th century, for being a "digital sweatshop", given that these services "[circumvent] a range of labor laws and practices, found in most developed countries, that govern worker protections, minimum wage, health and retirement benefits, child labor" (Zittrain, 2009, cited in MIT Technology Review, 2010). Authors such as Uddin (2012) and Cushing (2013) and studies by Horton and Chilton (2010) have stressed that "microworkers" typically work at a by far below average hourly wage (estimated to be less than USD 1.50). A survey of 200 workers on MTurk undertaken by Horton (2011) to investigate their perceived working conditions suggests, however, that "online workers view both offline and online employees more or less equally. In other words, they believe their chances of being treated fairly are as good or better online as they are offline" (MIT Technology Review, 2010). The study then suggests that regulation of the online labour market need to be carefully judged.

Since 2012, Amazon has embarked on an effort to verify all Amazon Payments accounts, including those of MTurk workers in light of criticism of declining working conditions of its international workers, but also due to risks of money laundering. This effort led to the deletion of many MTurk accounts (Ipeirotis, 2013). Furthermore, requesters are now restricted to entities based in the United States (Amazon, 2014a), and only workers in the United States and India can directly access the money transferred to their account, whereas other international workers can only receive the payment in the form of an Amazon gift card (Amazon, 2014b). As a result, MTurk workers, although an internationally diverse group of users, are mostly living in the United States and India today (Ipeirotis, 2010; Techlist, 2014), and the typical turker is not a person that completes tasks in a developing country for a living (Ross et al., 2010).

*Database administrator* is another traditional data specialist job at the early stage of the data value cycle. These data specialists use specialised software to store, organise and maintain data, such as financial information and customer shipping records. It is also their responsibility to make sure that data are available to users and also to secure databases and data warehouses from unauthorised access.[26] Database administrators hold well-established positions in firms, and the future outlook is positive: the share of jobs related to database administration has been growing since 1999 (see Figure 6.8) and the number of database administrators estimated to grow by 15% from 2012 to 2022 in the United States (BLS, 2014). This is four percentage points faster than the average growth across all occupations shown.

According to population surveys in the United States, the number of sectors employing one or more database administrators per 10 000 employees has increased over the past nine years (OECD, 2013a). In 2012, the five industries with the largest share of database administrators were financial activities (22 database administrators per 10 000 employees); professional and business services (12); wholesale and retail trade (6); manufacturing (6); and information (5, together with public administration and other services). The share of database administrators in these sectors has also increased significantly in recent years, with a remarkable peak in 2011.[27]

However, there is one trend that may reduce the need for database administrators in the near future despite the intensifying use of data and analytics across the economy, and that is the increasing use of online storage and analytics provided by cloud computing. Cloud computing has been described as "a service model for computing services based on a set of computing resources that can be accessed in a flexible, elastic, on-demand way with low management effort" (OECD, 2014c). It thus makes it less and less necessary to deploy internally managed databases, and as a consequence reduces the need for database administrators. As highlighted in OECD, 2010, cost savings through consolidation of ICT infrastructures is one of the expected benefits of cloud computing, and that includes savings of labour costs. Estimates focusing on software-as-a-service suggest that labour cost savings are the second biggest savings potential of cloud computing, after server software costs (see Voce et al., 2009 and MacManu, 2009).

*Actuaries, mathematicians, and statisticians* are data specialists that have in common that they use quantitative theories and methods to measure and analyse complex phenomena, including assessing uncertainty and dynamic processes to help businesses and clients develop strategies that maximise the business value under these uncertainties (BLS, 2014). These jobs often involve using statistical methods to collect and analyse data and help solve real-world problems in business, engineering, the sciences or other fields (BLS, 2014). Given the rapidly growing volume of data, these specialist jobs have been highlighted by many observers as the most promising job category in the near future. For example, Hal Varian, chief economist at Google, has been quoted saying that "the sexy job in the next 10 years will be statisticians. And I'm not kidding" (Varian cited by Lohr, 2009). This is supported by BLS (2014) estimates suggesting that mathematicians, actuaries and statisticians will be among the fastest growing occupations between 2012 and 2022 (26%). In the United States the overall share of mathematicians, actuaries and statisticians has been decreasing since 2012 (Figure 6.8), but their relative wages continue to increase (Figure 6.10). This suggests that the United States may be currently experiencing an undersupply of these specialists.

The importance of mathematicians, actuaries and statisticians has notably increased in the past year in several industries across OECD economies – particularly in insurance, advertising and market research. Mathematicians, actuaries and statisticians represented 1.5% of total employment in insurance in 2013, whereas in advertising and market research these occupations accounted for 0.9%, but the share has almost doubled compared to previous year (from 0.46%). Other industries show important increases in the levels of those occupations as well, although their intensities remain low – such as water collection, treatment and supply, where there was a fourfold increase in the level of intensity to 0.2% in 2013 since 2011.

A careful differentiation should be pointed out here, because employment prospects are not as positive for all mathematicians, actuaries and statisticians. At the 4th *OECD Global Forum on the Knowledge Economy* (GFKE, see Annex of Chapter 1 for the highlights), several participants noted that advanced analytic tools are now able to automate many simple tasks that statisticians used to do manually. Increasingly, these tools can, for instance, automatically fit thousands of statistical models to the available data and automatically generate and test different hypotheses (Davenport, 2014). A statistician relying on manual hypothesis testing can typically create only a few models per week. The increasing capacity of data analytics will most likely lead to less basic statistical work, which could lead to a negative employment outlook for less skilled mathematicians, actuaries and statisticians, including their related associate professionals and clerks (i.e. assistants).

### *Data scientists: The high-end all-round data specialists*

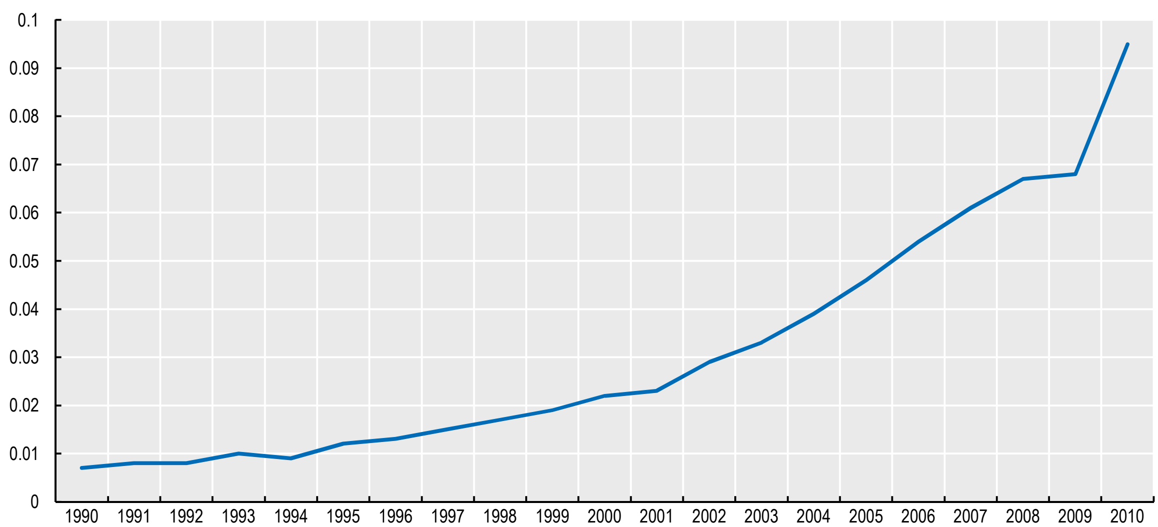
The increase in speed and variety of data-related activities along the data value cycle demands more efficient integration of these working activities, which – thanks to modern ICT tools, including data analytics and cloud computing – has become increasingly cost-effective to undertake. At the same time, business environments increasingly demand that data specialists be flexible and interdisciplinary since they require a great variety of skills, ranging from traditional computer science to mathematics and statistics to domain-specific skills and competence, including those related to business management, finance, marketing and health care, to name but a few. For many data specialists this means that specialist skills limited to one phase of the data value cycle will not be enough in the long run, and that there will be an increasing need to see the bigger picture to tackle all aspects of a problem, from initial data collection and data analysis to drawing conclusions for informing and even automating decision making.

In addition, the convergence of disciplines such as mathematics and statistics – but also, as mentioned above, natural and social sciences such as physics and economics – with computer science has been observed for some time (OECD, 2014d). As convergence has led to the cross-fertilisation of all related disciplines, it has also blurred the boundaries between the disciplines, including biology and computer science in the case of bio-informatics; finance, economics and computer science in the case of computational economics; and statistics and computer science – in particular artificial intelligence – in the case of machine learning, sometimes also referred to as statistical learning (see Chapter 3 of this volume). For (newly trained) data specialists, the convergence of disciplines provides an opportunity to respond to the increasing needs in current business and scientific environments for more flexible and interdisciplinary data-related work. But for many (established) data specialists this also increases the skills requirement considerably, as the use of new and more advanced data analytic tools and techniques is increasingly expected; many of these tools and techniques, such as advanced machine-learning algorithms, are being developed and used in computer science.

The trends presented above (convergence and the cost reduction in data-related activities) have led to the transformation of existing data specialist jobs into a job category or career path that is now commonly referred to as data scientist (Royster, 2013). The job title "data scientist" arose in recent years and has been most frequently used to describe professionals working on big data projects; however, there still is no widely agreed definition. The specific role of a data scientist is diverse and cannot be easily generalised, since a data scientist is expected to be an *all-round data specialist*. This has made measuring the number of data scientists particularly challenging; efforts to do so now rely on surveys such by King and Magoulas (2013, 2014) or analyses of social networks such as those of Patil (2011) and Tambe (2014). Analysis of job starters on LinkedIn by Patil (2011), for instance, shows that the number of persons embarking on careers as data scientists is growing exponentially, although the level remains low at 0.1% of all job starters in 2010 (Figure 6.15). Although they show a clear growing trend, these statistics have one serious limitation: they rely on workers to accurately classify themselves as "data scientists". In addition, Tambe (2014) points out that "there is a potential bias in online platform participation towards younger workers who use emerging technologies. Older IT workers using mature information technologies may have less incentive and lower proclivity to post their technical skills on LinkedIn".

Figure 6.15. **Growth of job starters listed in LinkedIn with a focus on data analytics and data science**

As percentage of all job starters




*Source*: Patil, 2011, based on LinkedIn.

Looking at various definitions used in the past years to characterise data scientists provides first insights into the occupations as well as the tasks and skills related to data scientists. Originally the term data scientists was coined by statistician Jeff Wu (1998, cited by Kuonen, 2014), who claimed that statisticians should be called data scientists since they spend most of their time manipulating and experimenting with data. In 2005, the US National Science Foundation (NSF) published a report that called for data scientists who are: "the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection" (NSF, 2005, Chapter 3). DJ Patil and Jeff Hammerbacher, who work with big data in LinkedIn and Facebook, respectively claim to have further developed the job title "data scientist" in 2008 in order to describe their position as "high-ranking professional with the training and curiosity to make discoveries in the world of big data" (Davenport and Patil, 2012).

The term, however, is somewhat vague. Many authors highlight the combination of statistics, programming and data visualisation skills as a key differentiating factor. Kenneth Cukier notes in *The Economist* (2010) that data scientists are a new kind of professional who combines the skills of software programmers, statisticians and storytellers/artists to extract insights from data. Similarly, Davenport (2012) calls data scientists "magicians who transform an inchoate mass of bits into a fit subject for analysis." In his opinion data scientists can extract data out of a server log, a telecom billing file, or the alternator on a locomotive, and analyse it. They also create new products and services for customers and also interact with senior executives and product managers (Davenport, 2012). According to the Data Science Association (2012), "data science" means the scientific study of the creation, validation and transformation of data to create meaning, and a "data scientist" is a professional who uses scientific methods to liberate and create meaning from raw data. Furthermore, they often work as data visualisers to create visualisations using both open source and proprietary tools to communicate their findings (UN Global Pulse, 2013).

In some cases, basic programming skills are not enough; what are needed are advanced (software) engineering skills, including expertise in machine learning (ML). Bertolucci (2014), Brave (2012) and the UK National Career Service (2014), for instance, stress that data scientists will often be in charge of integrating data from a variety of sources (i.e. data mashups) in order to develop data-driven applications building on ML algorithms. To do this data, scientists need to create and maintain complex software systems based on big data-specific technologies like Hadoop, Hive, Pig, HBase and Cassandra (Insight Data Engineering, 2014). Most of these technologies are so new however that few experts have sufficient knowledge or the expertise to work with them, and those with high levels of skills tend to concentrate in specific regions. Analysis of LinkedIn profiles by Tambe (2014) suggests that expertise in Hadoop, a major big data-related technology, is concentrated in certain regions in the United States, with the San Francisco Bay area being the most Hadoop-intensive region. His analysis underlines that geography matters for unleashing labour market spillovers, and provides an explanation for the systematic cross-regional firm-level variations in IT returns observed by many authors, such as Brynjolfsson and Hitt (2000), Dewan and Kraemer (2000), and Bloom and Van Reenen (2007). But these findings also call for a cautious interpretation of country-level employment statistics, which do not reflect (sub-)regional labour market concentrations and dynamics.

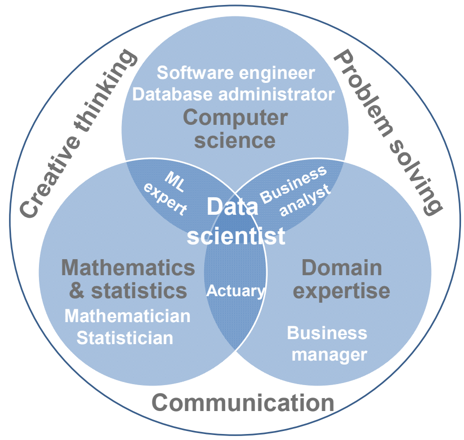
Last but not least, authors highlight that data scientists have domain-specific competence. Many authors stress in particular business-related skills and competence. Brave (2012), for instance, highlights that data scientists need to be able to analyse data sets to extract the domain- or business-relevant information. IBM (2014) stresses business acumen as a key ability of data scientists, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organisation approaches a business challenge. Brynjolfsson and McAfee (2012b) also highlight that data scientists have a high commercial awareness and knowledge of business processes to help decision makers reformulate their challenges so that big data can tackle them.

It should be noted that the need for domain-specific competence is not new, and has been often highlighted in the past with respect to ICT skills (see e.g. OECD, 2005 and 2012b). This also raises the question of the extent to which "data scientists" is a new job category. Some have criticised the current data scientist debate as overhyped, pointing to similar exaggeration around "big data". Harris, Murphy and Vaisman (2013), for instance, believe that the job title data scientists is more a buzzword then a real job description, since "the people doing this work used to come from more traditional and established fields: statistics, machine learning, databases, operations research, business intelligence, social or physical sciences, and more." In line with this, Kuonen (2014) believes that the occupation of a data scientist is not very different from what was formerly attributed to a statistician, since data have long played a role in advising and assisting operational and strategic thinking. For him as statistician, the term "data science", which is often used synonymously with data mining (Provost and Fawcett, 2013), is a part of statistics.

The advocates of data science, in contrast, claim that although statistics are an important part of data science, many of the key techniques for using big data are rarely taught in traditional statistics courses (Brynjolfsson and McAfee, 2012b). Kuonen (2014) also admits that there is a difference between data science and statistics. Traditional statistical analysis focuses on experimental data analysis and the testing of hypothesis, and thus takes a top-down approach; data science focuses on analysing observational data and aims at discovering new ideas with a bottom-up approach. Gartner researchers (Laney, 2012) found that data scientists are expected to work more in teams and to be more skilled at communication compared with traditional statisticians. They also frequently require experience in machine learning, computing and algorithms, and are required to have a PhD nearly twice as often as statisticians. Even the technology requirements for each role differ, with data scientist job descriptions more frequently mentioning Hadoop, Pig, Python and Java among others (Laney, 2012). Loukides (2011) summarises the discussion by explaining that the occupation of a data scientist is closely related to traditional occupations that require a strong mathematical background and computing skills. This is especially true since most data scientists on the job come from a discipline in which "survival" depends on getting the most from the data, such as physics, statistics and economics.

All these definitions and controversial debates presented above underline that data specialists have increasing skill requirements, with data scientists being – if not a "new" job category – at least the most advanced and talented data specialists. A transformation of data specialist jobs towards data scientists can also be observed. In other words, data specialists will increasingly need to combine skills and competences needed to collect, analyse, and use data across the data value cycle in a way that clearly creates value added for their organisation. In particular, data specialists will typically be required to have a mix of different skill sets, including computer science skills such as software engineering, database management, and machine learning (ML), as well as skills in statistics and domain-specific skills such as business management, marketing, finance and health (Figure 6.16). Data specialist skills therefore are not limited to (traditional) ICT specialist skills, although ICT specialist skills such as programming and database administration provide the basis for many future data specialist jobs including data scientists. In addition, "soft skills" such as communication, creative thinking and problem solving skills are also often increasingly highlighted as skill requirements (see discussion in previous section).

Figure 6.16. **Data specialist skills and competence mix**

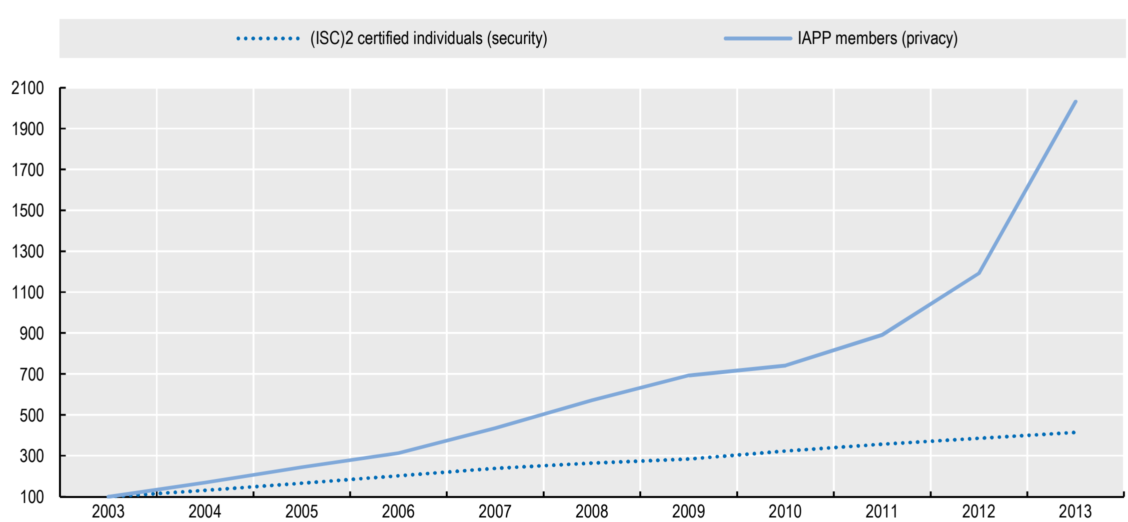



### *Security and privacy professionals*

For data-centred organisations, meeting privacy and security expectations requires more than legal compliance and sound security practices (see Chapter 5 of this volume). Under the 2013 revisions to the OECD Privacy Guidelines, for example, accountable organisations need to put in place multifaceted privacy management programmes, and stand ready to demonstrate them on request from a privacy enforcement authority (OECD, 2013c, paragraph 15). The range of skills needed to implement such programmes is broad, covering legal, technical, communications, governance and public relations aspects, for example. This has increased the need for experts in security and privacy. The steady growth in demand for security expertise seen over the past decade continues, while for privacy professionals the growth has been rapidly accelerating in recent years (Figure 6.17). For both security and privacy, the difficulty in finding available professionals with the required skills and expertise remains a challenge for organisations looking to strengthen their capacities in these areas.

Figure 6.17. **Trends in the number of certified/professional privacy and security experts, 2003-13**

Index, 100 = 2003




*Source:* Based on annual reports of the International Information Systems Security Certification Consortium [(ISC)[2], 2011] and of the International Association of Privacy Professionals (IAPP, 2013; 2014).

### *Cybersecurity professionals*

The number of cybersecurity professionals worldwide continues to rise steadily, as evidenced by the growing number of individuals with professional certifications for cybersecurity skills, such as the International Information Systems Security Certification Consortium [$(ISC)^2$] (Figure 6.17). As of the end of 2013, $(ISC)^2$ had certified 95 781 individuals worldwide, representing a fourfold increase over a decade. In the United States, the Bureau of Labor statistics shows that demand for graduate-level cybersecurity workers will rise by 37% over the next decade – more than twice the predicted rate of increase for the computer industry overall.[28]

Despite this rise, the supply of skilled cybersecurity professionals falls well short of demand. A 2013 report by Japan's National Information Security Center suggested a shortage of 80 000 information security engineers in the country (Humber and Reidy, 2014). Moreover, the report argued that most practicing cybersecurity professionals lack the skills necessary to effectively counter online threats.[29] Likewise, in the United Kingdom an analysis of official statistics on students leaving higher education in 2012-13 showed that less than 1% of computer science graduates in employment were in cybersecurity roles.[30] The National Audit Office of the United Kingdom has warned that it could take another 20 years to tackle the skills gap in trained cybersecurity staff.[31] The government has reacted to address this skills shortage. Specifically, the National Cyber Security Strategy of the United Kingdom aims to develop crosscutting knowledge, skills and capability. Through the National Cyber Security Programme, the Department for Business Innovation and Skills, the Government Communications Headquarters and the Cabinet Office have partnered to lead and support activity to increase cybersecurity skills at all levels of education.[32]

### *Privacy professionals*

One of the more important developments to improve the effectiveness of privacy protection measures has been the emergence of a professional class of privacy experts in organisations (Bamberger and Mulligan, 2011). In some cases there is a statutory basis to support or encourage the role of the privacy professional. For example, Germany's Bundesdatenschutzgesetz (Federal Data Protection Act) sets out specific requirements concerning data protection officials in organisations. Canada's federal private sector legislation, PIPEDA, requires an organisation to designate an individual(s) to be responsible for its personal data-handling activities. New Zealand's Privacy Act requires every agency in both the public and private sectors to appoint a privacy officer, and various pieces of US legislation require federal agencies to have Chief Privacy Officers or Senior Agency Officials for Privacy. Furthermore, the current EU Privacy Directive also contains a reference to a personal data protection official, and the proposed EU data protection regulation would require that data protection officers be appointed for all public authorities and for companies processing more than 5 000 data subjects within 12 months. This would further elevate the numbers of professionals.

The growth in the number of privacy experts has also been encouraged and supported by professional associations, setting the scene for the development of a privacy workforce, including chief privacy officers (CPOs) (Clearwater and Hughes, 2013). These associations provide training, certification, conferences, publications, professional resources and industry research to a growing membership. The largest and most global in reach – the International Association of Privacy Professionals (IAPP) – now has more than 18 000 members (a 24% increase from September 2013) in 83 countries around the world (Figure 6.17). Other associations include the Privacy Officers Network, through which senior privacy officers involved in the practical implementation of privacy initiatives meet and exchange ideas through a professional support network,[33] and national bodies like the Association Française des Correspondants à la Protection des Données à Caractère Personne (France)[34] and the Asociación Profesional Española de Privacidad (Spain).[35]

The steep growth in IAPP's membership numbers – from 10 000 members in 2012 to a projected 20 000 at the end of 2014 – demonstrates the broad recognition in the marketplace of the importance of sound data governance practices. In its Fortune 1000 Privacy Program Benchmarking Study, the IAPP documents that while budgets vary widely across Fortune 1000 companies, the average privacy budget is USD 2.4 million, 80% of which is spent internally on areas ranging from developing policies, training, certification and communications, to audits and data inventories. The study also highlights that privacy budgets are likely to grow, with nearly 40% of privacy professionals predicting an increase in their budget in the coming year (by an average of 34%) and 33% intending to hire new privacy staff. The IAPP's annual salary survey corroborates the results of the benchmarking study. The survey continues to demonstrate a steady increase in privacy officers' pay, with CPOs earning an average of USD 180 000 per year in the United States, and privacy leaders (who do not hold the title of CPO) earning an average of USD 131 000 in the United States, and USD 125 000 worldwide.[36]

Although the growth in security and privacy professionals documented in this section is both impressive and important, it does not fully capture the move to more deeply devote attention to these topics across workflows in some organisations. For these organisations, such issues are seen not just as the responsibility of designated privacy/security staff, but as a shared responsibility across the parts of the organisation that deal with personal data or matters impacting security. In particular, as companies move beyond viewing privacy as a compliance matter to be addressed by legal departments or as a technical issue handled by IT departments, they will need to put in place ethical review processes and ensure that they have privacy- and security-literate employees.

## 6.3. Promoting data-driven innovation and smoothing structural change

Seen against the background of one of the most important technological breakthroughs and innovation processes in human history, DDI carries great potential for improving the future of humankind in a world of global challenges, including unsolved development challenges. DDI is a new source of growth that can boost the productivity and competitiveness of all industries and economies. However, it is also disruptive, with a potential for "creative destruction" within labour markets that requires permanent and careful observation by policy makers. That effort is all the more necessary due to the potential of DDI to render many intellectually demanding jobs obsolete, including some data specialist jobs.

To what extent and within what time horizon DDI may lead to "technological unemployment" is still open and subject to academic debates.[37] But there is evidence that it may further increase inequality in earnings through skill-biased technological change, if not addressed by measures implemented through (inter alia) social and tax policies. In the current context of weak global recovery and lingering high unemployment in major advanced economies, the associated risks of unemployment and inequality in earnings thus deserves policy makers' attention.

The discussion in this section stems from the premise that "Resilient Economies and Inclusive Societies" is the yardstick to orientate policies for promoting jobs and "Environmentally Sustainable ('Greener') Growth", as underlined by Ministers and Representatives[38] in the OECD (2014a) Ministerial Council Statement. The following sections suggest that a "double strategy" is needed that i) supports the development and strengthening of the right mix of skills and competencies needed, including but not limited to data specialist skills, and ii) promotes social cohesion while addressing the risk of inequality that is rising in a number of countries (Herlyn et al., 2015). In the context of DDI, inequality could become a major issue, especially if access to high-quality education, which is urgently needed (before and after a first entry into the labour market) to take advantage of the job creation opportunities ahead, is limited to few (see Cingano, 2014).

### *Satisfying skill and competency needs*

Since the first Industrial Revolution, the education system has been key in supporting the need for structural adjustment in a "race between education and technology" (Goldin and Katz, 2008). For a long time and up until now, most societies have been successful in that race: new and better jobs for a better educated workforce have been the result, although structural adjustment often took time to take place and was not always very inclusive.

In light of the major implications of DDI discussed in this chapter, humans will need a broad education as a basis for the accelerating race ahead. An overly narrow education system, trimmed down to the specific jobs requirements of a particular time, will most likely not be the robust strategy needed to meet the skill challenges described in this chapter. The discussion presented here rather suggests that humans need to further develop their competitive advantage over machines and refine the skills that machines will not be able to perform at the upper end. This includes the development of a broader interdisciplinary understanding of multiple complex subjects, but also deeper insights into some domain-specific issues. A solid intellectual foundation in STEM (science, technology, engineering and mathematics), including in particular statistics and computer science, is necessary but not sufficient. At least as important is a true understanding of social and legal systems – in particular, of economics, ecology, human behaviour, legal requirements (e.g. relating to privacy and intellectual property rights), and – last but not least – ethics.

Creative thinking, problem solving, and communication skills have to be strengthened and cultivated as well as sensomotoric skills, as these will be the skills through which humans will outperform machines for a long time. In that respect humans are the best "combination" of abstract abilities and impressive sensomotoric skills, and this remains the key opportunity for job creation in the long run. But this also means that our "body and mind" need to be trained and kept in good shape throughout our lives. Achieving these huge education requirements will not be easy for individuals or for national education systems. Current pressure towards shorter time for education may make it more challenging to develop the required skills and competencies needed for a more inclusive "race between education and technology". In some cases formal education institutions may not be best placed to provide all the necessary skills. Therefore, efforts towards lifelong learning need to be supported by all stakeholders of the education system, including parents, formal education institutions, businesses, labour unions and governments. This requires a strategic approach for strengthening skill systems, as highlighted by the OECD Skills Strategy (OECD, 2012a) (see Box 6.7).

This section discusses means for developing the relevant data specialist skills, with a particular focus on formal education institutions. Further work would be required to fully apply the OECD Skills Strategy framework to assess how to better activate the supply of the skills needed for the data driven economy and how to put these skills to more effective use. Moreover, further reflection is needed to better understand how to better develop the competitive advantages humans have over machines.

Box 6.7. **The OECD Skills Strategy**

The OECD (2012) Skills Strategy framework provides countries with a strategic approach to strengthen their skills system in building, maintaining and using their human capital to boost employment and economic growth, and promote social inclusion and participation. It encompasses the following objectives:

1. Developing relevant skills, by i) encouraging and enabling people to acquire the right skills throughout life, ii) fostering international mobility of skilled people to fill skill gaps, and iii) promoting cross-border skills policies
2. Activating skills supply, by i) encouraging people to offer their skills to the labour market and ii) retaining skilled people in the labour market
3. Putting skills to effective use, by i) creating a better match between people's skills and the requirements of their job and ii) increasing the demand for skills.

*Source:* OECD, 2012a.

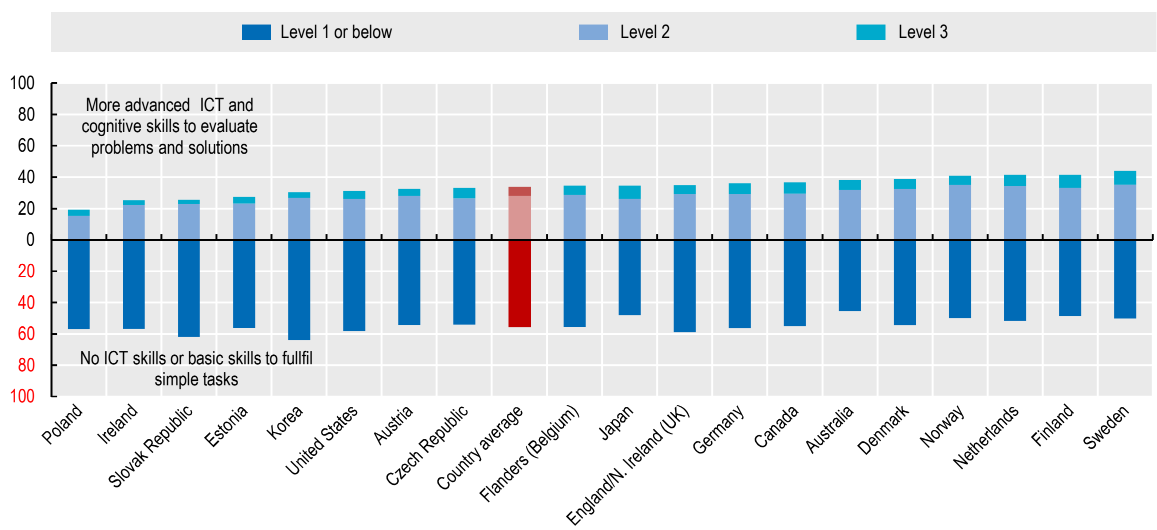
### *Developing the basic skills needed*

In the past there have been considerable mismatches between the supply of and demand for ICT skills in general and software skills in particular (OECD, 2012a). Shortfalls in domestic supply, owing to a large share of students leaving compulsory education; lack of educational courses and little vocational education or on-the-job training in industry; and the low share of female employees in ICT specialist occupations were often highlighted as factors limiting the availability of ICT specialist skills. And this could remain true for data specialist skills. Furthermore, restrictions on the immigration of highly skilled personnel and difficulties in international sourcing of analytical tasks – which could intensify due to current considerations of data localisation requirements – put further pressure on national education systems to develop the right mix of skills needed for the data-driven economy.

Besides language and reasoning skills, creativity and social intelligence, and perception and sensomotoric skills, basic ICT literacy needs to be much further developed, given that it has become the skill foundation of the workforce in the digital economy; this includes knowledge about security and privacy risks, as specifically identified in the revised OECD Privacy Guidelines' call for "complementary measures". However, an OECD (2014g) study based on data from the OECD's Programme for the International Assessment of Adult Competencies (PIAAC) reveals that 7% to as much as 27% of adults have no experience in using computers, or lack the most elementary computer skills, such as the ability to use a mouse. The study also shows that in OECD countries, only 6% of the population is categorised with the "highest level" of ICT skills, meaning "they can complete tasks involving multiple applications, a large number of steps, impasses, and the discovery and use of *ad hoc* commands in a novel environment". In countries such as Austria, the United States, Korea, Estonia, the Slovak Republic, Ireland and Poland, the share is 5% and below (Figure 6.18). This suggests that for most OECD countries, the basis for developing data specialist skills is very weak.

Figure 6.18. **Level of proficiency in problem solving in technology-rich environments, 2012**

As a percentage of 16-65 year-olds



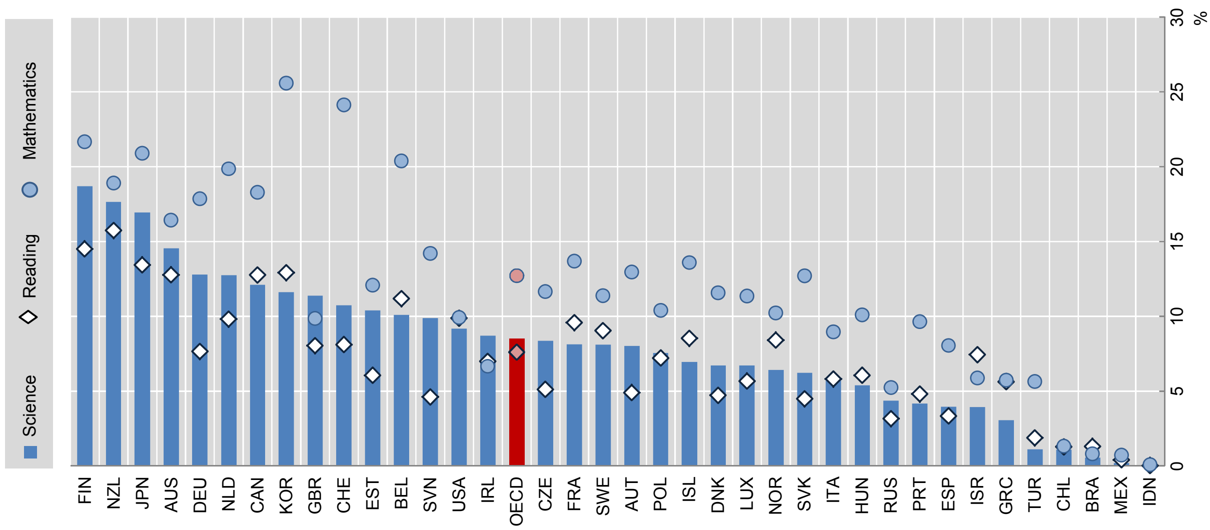

*Note:* Problem solving in technology-rich environments requires "computer literacy" skills (i.e. the capacity to use ICT tools and applications) and the cognitive skills required to solve problems. Level 1 or below possesses no ICT or basic skills to fulfil simple tasks; levels 2 and 3 require more advanced ICT and cognitive skills to evaluate and find solutions.

*Source*: *OECD Science, Technology and Industry Outlook 2014*, based on OECD's Programme for the International Assessment of Adult Competencies (PIAAC), http://dx.doi.org/10.1787/888933151932.

The discussion presented in this chapter highlighted mathematics proficiency as an important foundation for data specialist skills besides language and reasoning skills. Mathematics and statistics prepare students to work with data analytics. "Math helps students develop the logical thinking and problem-solving skills they need. Statistics provides the analytical knowledge that they need to properly study the data and to interpret the results in a meaningful way" (Royster, 2013). That means that countries with a high share of top-performing students in mathematics but also reading and science are more likely to develop talent pools for future data specialists. Results from the 2009 OECD Programme for International Student Assessment (PISA) on the science, reading and mathematics proficiency of 15-year-olds show that 13% of students in the OECD area were top performers in mathematics, 9% in science, and 8% in reading (Figure 6.19). The share of top performers in mathematics is highest in Korea, Switzerland, Finland, Japan and Belgium. It is the lowest in Mexico, Chile, Turkey, Greece and Israel, besides partner economies Indonesia, Brazil and the Russian Federation. These low performers could be facing difficulties in developing data specialist skills the next five to ten years.[39]

Figure 6.19. **Science, reading and mathematics proficiency at age 15, 2009**

Percentage of top performers



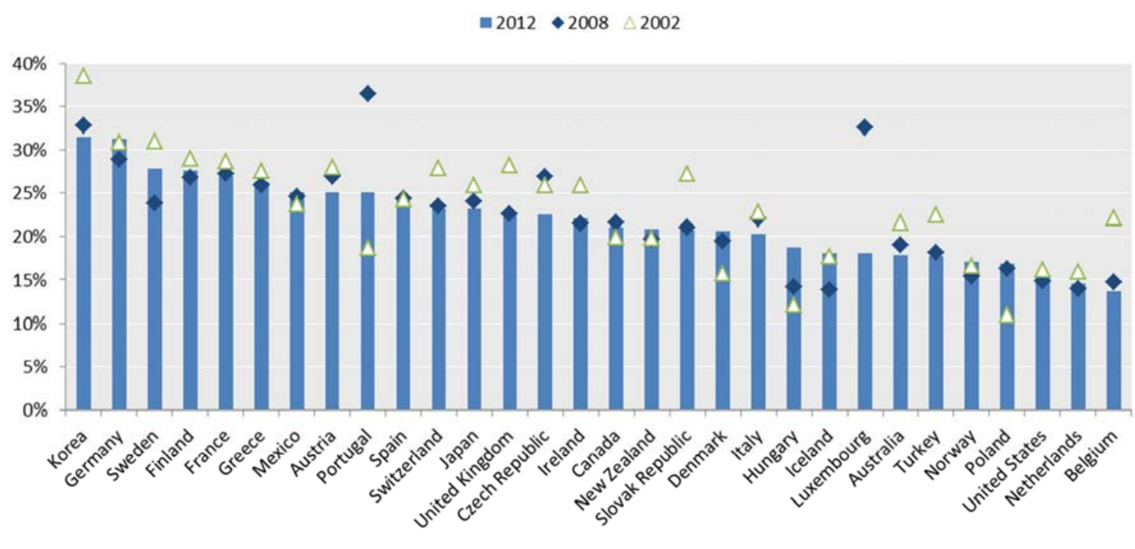

*Source:* OECD Science, Technology and Industry Scoreboard 2013, http://dx.doi.org/10.1787/888932890675.

## *Developing the higher skills needed*

Higher education institutions have a pivotal role to play in providing the higher-level skills needed by data specialists. This is especially the case for computer science skills, since very few secondary education institutions provide skills such as programming and database management. These skills are essential however, since they not only provide the ability to develop, maintain, and operate data-driven systems, but they also train logical thinking and problem solving (Royster, 2013). Evidence suggests that the supply of computer science graduates is progressing. Over a period of five years (2000-05), the number of computer science graduates in OECD countries for which data are available almost doubled, but then started to decline until 2010. Since then, the number of graduates in computer science has been increasing, reaching almost the same level in 2012 as in 2006.

But in addition to a degree in computer science (often with an emphasis on data management, data mining or artificial intelligence), a significant number of data specialists have a degree in experimental physics, molecular biology, or bio-informatics, which often involves analysis of large data sets (Loukides, 2011; Rogers, 2012). This suggests that data specialist skills are not only provided in computer science study programmes, but they are also more likely to become an integral part of science, technology, engineering and mathematics (STEM) disciplines, in particular given the convergence of the disciplines with computer science as highlighted above and the increasing data intensity of science and engineering (see Chapter 7 of this volume). However, the share of all STEM graduates from OECD countries for which data are available declined from 22% in 2000 to 21% in 2012 (see Figure 6.20), and the participation of women in tertiary-level graduate STEM programmes remains low (see Box 6.5). These trends indicate a long-term stagnation of the relative supply from high-demand science and technology oriented fields. In many countries in 2012, the supply of higher education graduates from STEM-related study fields indeed stagnated (Norway, Ireland, Poland and Greece) or even declined (Czech Republic, Korea, France, Slovak Republic, Spain and Italy) in absolute numbers. But the decline is also due to a faster growth in the number of graduates in non-STEM fields (see e.g. Austria).

Figure 6.20. **STEM (science, technology, engineering and mathematics) graduates**
As percentage of total graduates



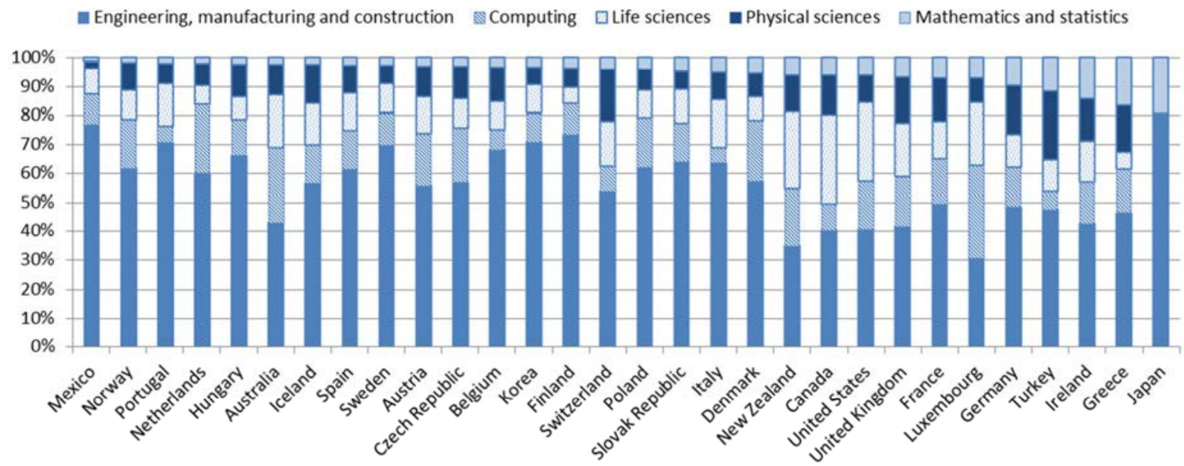

*Note:* Graduates are those who successfully complete an educational programme during the reference year of the data collection. The year shown is the year in which students graduate, with the exceptions of Denmark, Finland, France (until 2002) and Italy where students graduate the previous year.

*Source:* Based on UNESCO-OECD-Eurostat (UOE) data collection on education statistics, compiled on the basis of national administrative sources, reported by Ministries of Education or National Statistical Offices.

A closer look at STEM graduates by disciplines is worth taking, in particular the share of mathematics and statistics graduates. Not only are there huge cross-country variations, but when considering that mathematics and statistics are most likely the most demanding data specialist occupations, with indications of shortages in particular in the United States, it is striking to see these disciplines account by far for the lowest share of STEM graduates in some countries (between 2% in countries such as Mexico, Norway, Portugal, and the Netherlands, and 10% in Germany, 12% in Turkey, 14% in Ireland, and 17% in Greece) (Figure 6.21).

At the same time, STEM should not be overrated, since they are not the only source for the skills and competences needed in a data-driven economy. As highlighted above, a solid intellectual foundation in STEM is not sufficient. But even when focusing on data specialist skills, one can observe that these skills may even be acquired in disciplines beyond STEM. For example, the emergence of trends such as “data journalism”, where journalism is centred on the use data and data visualisation, suggests that data analytic skills may also be part of tomorrow’s curricula for a wider range of study programmes including journalism.

Figure 6.21. **STEM graduates by disciplines, 2012**




*Source*: Based on UNESCO-OECD-Eurostat (UOE) data collection on education statistics, compiled on the basis of national administrative sources, reported by Ministries of Education or National Statistical Offices.

In order to meet the increasing demand for data specialists, an increasing number of education institutions are offering specialised master's-level programmes related to data and analytics, some of which are advertised as "data science" programmes (Violino, 2014). In the United States, a growing number of data science programmes, often master's-level programmes, can be observed. Northwestern and North Carolina State University offer, for instance, a Master of Science in Analytics (MSiA) programme that provides students with "coursework in statistics, modeling, operations research, quantitative analysis, decision analysis, databases, and data management form the core of the curriculum".[40] Most programmes offer a similar curriculum. Few however provide domain-specific skills and competence such as business and entrepreneurial, and communication skills in addition. And even less provide coursework on privacy protection and cybersecurity.

The very new and rapidly changing nature of ICTs including data analytics makes workplace training, in addition to formal education, increasingly important for augmenting and adapting workers' skills. As Royster (2013) confirms, "even workers who have a statistics or data analysis background need to stay current with the fast-changing world of big data". This is especially true for older workers, for whom skills acquired through the educational system are likely to be missing or substantially depreciated in the field of data specialist skills. Evidence in the past suggested that the share of firms with vocational training on ICTs decreased between 1999 and 2005, which could raise concern in policy makers. Although no representative statistics exist, there is anecdotal evidence in the case of data analytics that in particular large firms are offering data analytic training for workers who have little formal training. Many data scientists training programmes developed by private sector companies such as IBM are also offered to the public (Marsan, 2012). In addition, Massive Open Online Courses (MOOCs) on data analytics are increasingly being offered. For instance, in February 2015 Coursera, one of the leading MOOC providers, offered 69 courses related to "Statistics and Data Analysis". This was by far the most frequent course category in Coursera, before "Information, Tech & Design" (25), "Social Sciences" (15) and "Mathematics" (14).

## *Addressing the risks of rising income inequality*

Beside policies aiming at enhancing skills, further policy consideration may be needed to smooth structural change and in particular address the inequality that could increase due to DDI. A recent report by Cingano (2014) provides evidence suggesting that "reducing income inequality would boost economic growth" (OECD, 2014e). The study suggests that economies where income inequality is decreasing grow faster than those with rising inequality. The study underlines that lack of investment in education is the main mechanism through which growth is hampered due to inequality: in particular for children from poor socio-economic backgrounds, lack of investment in education lowers social mobility and hampers their skills development, thereby negatively affecting growth potentials in the long run. The study therefore highlights the need "to promote equality of opportunity in access to and quality of education" (Cingano, 2014). But the study also stresses that "redistribution policies via taxes and transfers are a key tool to ensure the benefits of growth are more broadly distributed and the results suggest they need not be expected to undermine growth" (Cingano, 2014).

As highlighted above, DDI offers potentially significant increases in labour productivity through the automation of cognitive and manual tasks, many of which are high value added tasks considered to be not susceptible to machines automation (i.e. taxi drivers affected by driverless cars; research clerks and assistants affected by expert systems such as IBM Watson). Therefore, several authors including Ford (2009), Cowen (2013), Cukier and Mayer-Schönberger (2013), Frey and Osborne (2013), Levy and Murnane (2013), Brynjolfsson and McAfee (2014), Rifkin (2014) and Elliott (2014), have highlighted the potential negative implications of data-driven automation on wage and income inequalities.[41] Brynjolfsson and McAfee (2012a; 2014), for instance, refer to work by Acemoglu and Autor (2010) on "skill-biased technical change" to present evidence on the growing relative wages of higher skilled workers compared to lower skilled workers. This trend could intensify with DDI, which tends to amplify employment polarisation as middle income jobs become more susceptible to machine automation as highlighted above.

Some of these implications were highlighted already in the OECD (2013b) Report on *Supporting Investment in Knowledge Capital, Growth and Innovation*, which looked at the policy implications of knowledge-based capital (KBC), and concluded that:

> *KBC-based economies rewards skills and those who perform non-routine manual and cognitive tasks, but may also reward investors (who ultimately own much of the KBC) over workers (in the United States, for instance, wages as a share of GDP are at an all-time low). Rising investment in KBC can create winner-takes-all opportunities for a few, while entire occupational categories can be replaced by machines and software. (OECD, 2013b)*

OECD (2013b) points to a much broader set of issues about the remuneration of labour versus remuneration of capital,[42] that many of the authors listed above have also called attention to in light of the still open academic debate about "technological unemployment". Their main argument can be summarised as follows: if the need to sufficiently pay a highly educated workforce should be eliminated via smart applications that can (partly) perform knowledge- and labour-intensive tasks with less labour, including less educated and paid labour, then the known pattern of income distribution could be in danger. Often, redistribution policies including also (i) negative income tax (NIT) and (ii) an unconditional basic income, are then proposed for consideration. NIT, such as proposed for example by Friedman (1962), provides citizens whose income is below a certain threshold with supplemental payments from the government (instead of paying taxes to the government). Some have suggested that earned income tax credits (EITC) already provided for example in the United States to low or moderate income working individuals would be a policy measure similar to NIT (Farrell, 2013). Unconditional basic income, in contrast, is paid to every citizen either employed or unemployed. It is sometimes suggested that its amount should reflect national threshold definitions under which citizens would be categorised as "poor". The impact of both policy instruments and the challenges they raise are still subject to controversial debates, as is the question about "technological unemployment". That said, inequality will remain high on policy makers' agenda, in particular if the availability of natural resources remains a huge constraining factor.

## 6.4. Key findings and policy conclusions

DDI can be disruptive with the potential to amplify employment polarisation as it affects a broader range of middle income jobs – the segment of the population that glues our societies together. Moreover, DDI can contribute to employment polarisation through (i) data-driven decision automation affecting white collar jobs, and (ii) the enabled new generation of autonomous machines affecting blue collar jobs in particular in manufacturing and logistics. DDI may even negatively affect some data specialist jobs such as data entry clerks, database administrators, and statisticians.

That said, it is important to stress that DDI should not be seen in isolation from the rest of the economy. In particular, the dynamics induced by DDI need to be carefully studied. As technological change tends to increase productivity, it reduces costs and increases demand for goods and services, which in turn can help generate more jobs in other parts of the economy. And this is also true for the new goods and services that DDI directly enables. Further studies taking into account these dynamics are therefore needed in order to better understand the employment effects of DDI.

What this chapter clearly showed however is that further investments in education are needed in order to promote the adoption of DDI across society, and to support the developments of the right mix of skills and competence needed for the (re-)creation of decent jobs. The chapter pointed to current debates about the risks that DDI could worsen wage and income inequalities as it amplifies employment polarisation. Evidence shows that economies with decreasing income inequality have grown faster than those with rising inequality (Cingano, 2014). Supporting the development and strengthening of the right mix of skills and competencies, while promoting social cohesion and addressing the rising inequality, could be part of a "double strategy" to prepare societies for the future (Herlyn et al., 2015).

The chapter presented evidence on the data intensification of OECD economies. Data specialist skills are increasingly in demand across industries, although the share in total employment tends to remain low. However, there are signs of an insufficient supply of (basic) skills related to ICT and STEM more broadly, with indications of actual skills shortages with respect to statisticians, mathematicians and actuaries. Developing mathematic and statistic proficiency is essential to appropriately use data and analytics, and how to deal with their limitations discussed in Chapter 3 of this volume.

But the implications of DDI for education are far-reaching, going beyond skills related to ICTs and even STEM. The discussion presented in this chapter suggested that education systems should support a broader interdisciplinary understanding of multiple complex subjects but also deeper insights into some domain-specific issues. This calls for the development of a solid intellectual foundation in STEM *in combination* with a sufficient understanding of human behaviour and social systems such as provided in humanities. This combination would help enhance qualitative reasoning in addition to quantitative reasoning to enhance the sense of responsibility of future data-informed decision makers.

Furthermore the chapter highlighted that soft skills such as (i) creativity, (ii) problem solving and (iii) communication skills are key for ensuring employment in a data-driven economy in the long run. Furthermore, (iv) highly developed sensomotoric skills will also become a key competitive advantage of humans over machines. These skills, if cultivated with the support of education systems and accompanied by political attention and good co-operative global governance, may lessen concerns related to technological unemployment. This will be more the case if individuals can enhance and complement their talents to use technology to “dance” with the machines instead of “racing” against them.

## Annex – Selected statistical definitions of data specialist occupations

Table 6.A1 **Europe: Occupations included in the operational definition of the Data specialists.**

Based on ISCO-08 (3-digits)

| ISCO-08 | EN Title |
|---|---|
| **212** | Mathematicians, actuaries and statisticians |
| **252** | Database and network professionals |

Table 6.A2 **United States: Occupations included in the operational definition of data specialist**

Based on SOC 2010 (6-digits)

| SOC 2010 | EN Title | Corresponds to ISCO-08 codes (4-digits) | |
|---|---|---|---|
| **15-2011** | Actuaries | 2120 | Mathematicians, actuaries and statisticians |
| **15-2021** | Mathematicians | 2120 | Mathematicians, actuaries and statisticians |
| **15-2031** | Operations Research Analysts | 2120 | Mathematicians, actuaries and statisticians |
| **15-2041** | Statisticians | 2120 | Mathematicians, actuaries and statisticians |
| **19-3022** | Survey Researchers | 2120 | Mathematicians, actuaries and statisticians |
| **15-1141** | Database Administrators | 2521 | Database designers and administrators |
| **15-1142** | Network and Computer Systems Administrators | 2522 | Systems administrators and |
| **15-1143** | Computer Network Architects | 3514 | Web technicians |
| **15-1122** | Information Security Analysts | 2523 | Computer network professionals |
| **15-1199** | Computer Occupations, All Other | 2529 | Database and network professionals not elsewhere classified |
| | | 2529 | Database and network professionals not elsewhere classified |
| | | 2519 | Software and applications developers and analysts not elsewhere classified |

Table 6.A3 **Australia: Occupations included in the operational definition of data specialist**

Based on ANZSCO 2010 (6-digits)

| ANZSCO version 1.2 | | Corresponds to ISCO-08 codes (4-digits) | |
|---|---|---|---|
| 224111 | Actuary | 2120 | Mathematicians, actuaries and statisticians |
| 224112 | Mathematician | | |
| 224113 | Statistician | | |
| 262111 | Database Administrator | 2521 | Database designers and administrators |
| 262113 | Systems Administrator | 2522 | Systems administrators |
| 263112 | Network Administrator | | |
| 263111 | Computer Network and Systems Engineer | 2523 | Computer network professionals |
| 263113 | Network Analyst | | |
| 262112 | ICT Security Specialist | 2529 | Database and network professionals not elsewhere classified |

Table 6. A4 **Canada: Occupations included in the operational definition of data specialist**

Based on NOC 2011 (4-digits)

| NOC 2011 | | Corresponds to ISCO-08 codes (4-digits) | |
|---|---|---|---|
| 2161 | Mathematicians, Statisticians and Actuaries | 2120 | Mathematicians, actuaries, statisticians |
| 2172 | Database analysts and data administrators | 2521 | Database designers and administrators |

# Notes

1 For instance, it is unclear whether those firms adopting DDI became more productive due to DDI-related investments, or whether they were more productive in the first place. Furthermore, these studies rarely control for the possibility that some firms may have eventually seen a reduction in their productivity due to DDI, and as a result may have discontinued their investments in it.

2 For consumer goods and retail firms it is the single biggest barrier, cited by two-thirds of respondents from those sectors. Furthermore, MGI (2011) estimates that the demand for deep analytical positions in the United States could exceed supply by 140 000 to 190 000 positions by 2018.

3 This chapter benefited from, and partly builds on, reflections on the OECD project on data-driven innovation by Estelle L.A. Herlyn, Thomas Kämpke, Franz Josef Radermacher, and Dirk Solte (Research Institute for Applied Knowledge Processing, Ulm, Germany, FAW/n; see Herlyn et al., 2015).

4 With information and communication technologies (ICTs), we have seen the highest innovation speed and greatest penetration rate of new technologies, ever. At the heart of this development is the extreme speed in cost reduction as described by Moore's Law, which holds that processing power doubles about every 18 months relative to the cost or size of central processing units (CPUs) (Moore, 1965). In other words, for decades mankind has been enhancing the performance of processors by a factor of 10 000 every 20 years, which means an improvement factor of more than a trillion over the past 60 years, since the time work on the first transistors or chips started. These are huge achievements – one could easily argue that there has never been so much change induced by technology in such a short time.

The main reason for this explosion of improvement is the possibility of miniaturizing or compressing information. That means that encoding of one unit of information (one bit) requires ever less physical space. This is because the coupling of information and its physical manifestation is very loose. We can make the encoded information (e.g. numbers) always smaller, without changing the results of subsequent algorithmic computations on the information, be it e.g. arithmetic or Boolean operations. That means that in order to add numbers, the size of the physical representation of the numbers is not of principal importance, in contrast to e.g. to physical good such as a car, where the size of the car is essentially a given and not a variable.

5 One way of measuring the IoT is by looking at the number of SIM-cards and phone numbers allocated to M2M communication devices on mobile networks (OECD, 2015). The data show in many countries the market is growing briskly. Most countries report double-digit growth between 2012 and 2013, though most countries do not have data for 2011, so it is hard to analyse trends. Some operators are also reporting on the number of connected devices. AT&T in the United States, for example, reports that it connected 1.3 million devices in the second quarter of 2014, of which 500 000 were vehicles.

6 To make this work, machine learning uses many techniques, also used in data analytics, such as statistical and regression analysis; (unsupervised) learning algorithms and cluster analysis.

7 See the trends referred to as "smart manufacturing", the Industrial Internet, or Industry 4.0.

8 Similar systems underlying ATS are being provided by start-up companies offering online financial advice services at much lower fees than traditional financial advisors (Coombes, 2013).

9 See www.youtube.com/watch?v=sOLXOsiX0Qk on "current and future applications of machine-learning within law, and … automation in the context of legal tasks currently performed by attorneys, including predicting the outcomes of legal cases, finding hidden relationships in legal documents and data, electronic discovery, and the automated organization of documents", accessed 18 May 2015.

10 For example, workers in Amazon's warehouses in the United Kingdom are reported to walk between 11 and 24 kilometres per day (O'Connor, 2013).

11 Before the system can function, it has to model the position of all goods in the warehouse and the most efficient paths and distribution.

12 Unlike Amazon, it cannot move the shelves because the products are bulky and heavy. Instead, it has created a three dimensional storage facility.

13 *Source:* Video clip of Symbotic presentation at www.symbotic.com/robots-in-the-warehouse-expanding-beyond-manufacturing/, accessed 18 May 2015.

14 In this scenario packages are not on pallets, but stacked individually into a delivery van. This again was a computationally hard problem in the past, because it required the packages to be loaded in reverse order from being delivered, but in such a way that the truck can hold the most packages securely. Furthermore, it required three dimensional perception to correctly identify and handle the packages, which are of different sizes and weights.

15 The Amazon warehouse example shows that workers are still necessary to fill orders, to actually pick the products, and to fill the boxes.

16 One sector where autonomous machines are increasingly a reality is the agriculture sector. There are a great number of examples. Algorithms and robots already sort plants such as orchids into various classes and groups, based on optical recognition. Robots harvest lettuce and recognise rotten apples. The spraying of fields is done by tractors that steer themselves and only need minimal operator intervention. Combine harvesters can operate semi-autonomously or work together with a lead harvester. Algorithms vary the spraying of pesticide and fertiliser based on yield data from previous years. The robotic tractor cannot be far off.

17 Together with the MIT economist David Autor, they have examined the changes in occupational distribution in the United States by categorising the work in five areas: 1) solving unstructured problems, 2) working with new information, 3) routing cognitive tasks, 4) routine manual tasks and 5) non-routine manual tasks. The result shows a clear trend, as described in Figure 6.3.

18 It is interesting to note that, given the impressive manifestations of machine intelligence, humans may need to further "re-discover" the importance of their bodies.

19 But he adds that, "the future will bring us The Unaccountable Freestyle Team, The Scary Freestyle Team, and The Crippled Freestyle Team, all at once" (Cowen, 2013, p. 131).

20 Estimates are based on the voluntary, ad hoc module in the EU Community Innovation Survey 2010 on the skills available in enterprises and on methods to stimulate new ideas and creativity. The indicator corresponds to the percentage of firms in the relevant innovation category responding affirmatively to the question: "During the three years 2008 to 2010, did your enterprise employ individuals in-house with the following skills, or obtain these skills from external sources?" Innovative enterprises had innovation activities during 2008-10, relating to the introduction of new products, processes, and organisational or marketing methods. This includes enterprises with ongoing and abandoned activities for product and process innovation. The question on innovation-relevant skills also applies to non-innovative enterprises. Estimates are based on firms with "core" NACE Rev. 2 economic activities (B, C, D, E, G46, H, J58, J61, J62, J63, K and M71).

21 In these countries, the share of data specialists has decreased considerably in recent years. It is important to note that in all countries aside from Greece and Turkey, data specialists mainly comprise database and network professionals.

22 In 2012, information and communication industries accounted for 3.6% of total employment in the OECD area. In nearly all countries, IT and other information services are the largest component of the information and communication industries, accounting for 40%, i.e. 1.4% of total employment in the OECD area. With above 2% of total employment, Ireland was the country with the largest share (2.7%), followed by Luxembourg, the United Kingdom, Finland, Sweden and Denmark. Australia, Greece and Mexico had among the lowest shares, all equal or below 0.5%. Since 2004, the share of IT and other information services in total employment has grown continuously, going from 1.2% to 1.4%. IT and other information services over the last decade have been resilient and driving employment growth, especially during the recent financial crisis where the employment losses were less important than in other industries, with a growth rate in 2009 of less than -1% while the information industries and total employment growth rates were respectively of -3.5% and -2.5%. However, since 2010 IT and information services have been growing quite fast which suggests that ICTs and in particular, IT and information services are playing a significant role in the upcoming recovery.

23 According to data published by the World Information Technology and Services Alliance (WITSA), telecommunications (11.5%), financial services (6.6%), transport (5.1%), health care (4.1%) and government (3.8%) are the five most ICT-intensive sectors. Using ICT intensity as a proxy for data intensity assumes that data-intensive industries have higher ICT expenditure than industries with low data intensity. That assumption can be easily challenged, since data analytics requires less investment in ICTs today (because of cloud computing). In an historical perspective however, this approach can still be useful.

24 Typical examples include text-heavy data sets such as text documents and e-mails, as well as multimedia content such as videos, images and audio streams. Unstructured data account for the largest share of the global data volume by far. According to some estimates, not even 5% of the digital universe can be considered structured or semi-structured data.

25 For example, optical character recognition (OCR) can transform images of text into machine-encoded text, which then can be further processed and used for data analytics, in particular natural language processing (NLP), for tagging or for extracting relevant patterns.

26 Database administrators sometimes share these tasks with network and computer systems administrators, computer network architects, and information security analysts.

27 In 2011, financial activities, professional and business services, information, and public administration were the sectors mainly contributing to the increase in share of database administrators in the United States.

28 See www.bbc.com/news/business-26647795

29 See www.businessweek.com/articles/2014-07-24/proposed-law-would-fix-japans-lax-cybersecurity

30 See www.ft.com/intl/cms/s/0/76b1eef4-1d3c-11e4-8b03-00144feabdc0.html

31 See www.bbc.com/news/business-26647795

32 See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/289806/bis-14-647-cyber-security-skills-business-perspectives-and-governments-next-steps.pdf

33 See www.privacylaws.com/Privacy-Officers-Network/

34 See www.afcdp.net/

35 See www.apep.es/

36 See 2013 IAPP Privacy Professionals Role, Function and Salary Survey, https://privacyassociation.org/resources/article/2013-iapp-privacy-professionals-role-function-and-salary-survey

37 See Ford (2009), Wilkinson and Pickett (2010), Randers (2012), Cowen (2013), Cukier and Mayer-Schönberger (2013), Frey and Osborne (2013), Levy and Murnane (2013), Brynjolfsson and McAfee (2014), Herlyn and Radermacher (2014), Piketty (2014), and Rifkin (2014).

38 These include ministers from and representatives of Australia, Austria, Belgium, Canada, Chile, Colombia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, the United States and the European Union.

39 It is worth noting at this point that new ways to teach math are being developed; see for example Wing, 2008.

40 www.analytics.northwestern.edu/curriculum-and-career-prospects/Course_description.html, accessed 18 May 2015.

41 This leads directly to the question of the implications on DDI on wage and income inequalities, and to the question of the right kind of balance in income concerning the so-called "efficient inequality range" (Cornia and Court, 2001).

42 Cowen (2013, pp. 38-40), for instance, observes that: "In 1990, 63 percent of American national income took the form of payments for labour, but by the middle of 2011 it had fallen to 58 percent." And this trend is not limited to the United States; many OECD countries including France, Germany, and Japan have seen similar trends as highlighted by Piketty (2014). Other authors such as Atkinson (1975), Wilkinson and Pickett (2010), Herlyn (2012), and Herlyn and Radermacher (2014) have discussed the implications more broadly, highlighting for example a trend towards "precarisation" or "neo-feudalisation".

## *References*


Abedi, V. et al. (2012), "An automated framework for hypotheses generation using literature", *BioData mining*, Vol. 5, No. 1.

Acemoglu, D. and D. Autor (2010), "Skills, tasks and technologies: Implications for employment and earnings", *NBER Working Paper* No. 16082, National Bureau of Economic Research.

Amazon (2014a), Amazon Mechanical Turk FAQs, http://aws.amazon.com/mturk/faqs/#Can_international_Requesters_use_Amazon_Mechanical_Turk_to_get_tasks_completed, accessed 17 December 2014.

Amazon (2014b), Worker Web Site FAQs, https://www.mturk.com/mturk/help?helpPage=worker#how_paid, accessed 17 December 2014.

Andreessen, M. (2011), "Why software is eating the world", *The Wall Street Journal*, 20 August, http://online.wsj.com/article/SB10001424053111903480904576512250915629460.html.

Atkinson, A.B. (1975), *Economics of Inequality*, Oxford University Press.

Aul, W.R. (1972), "Herman Hollerith: Data processing pioneer", *Think*, IBM's employee publication, November, pp. 22-24, www-03.ibm.com/ibm/history/exhibits/builders/builders_hollerith.html, accessed 25 May 2015.

Autor, D.H. and D. Dorn (2013), "The growth of low-skill service jobs and the polarization of the US labor market", *The American Economic Review*, Vol. 103, No. 5, pp. 1553-97, http://economics.mit.edu/files/1474, accessed 25 May 2015.

Bakhshi, H., A. Bravo-Biosca and J. Mateos-Garcia (2014), "Inside the datavores: Estimating the effect of data and online analytics on firm performance", Nesta, March, www.nesta.org.uk/sites/default/files/inside_the_datavores_technical_report.pdf, accessed 25 May 2015.

Bakhshi, H. and J. Mateos-Garcia (2012), "Rise of the datavores: How UK businesses analyse and use online data", Nesta, November, www.nesta.org.uk/sites/default/files/rise_of_the_datavores.pdf, accessed 13 May 2015.

Bamberger, K.A. and D.K. Mulligan (2011), "Privacy on the books and on the ground", *Stanford Law Review*, Vol. 63.

Barua, A., D. Mani, R. Mukherjee (2013), „Impacts of effective data on business innovation and growth", Chapter two of a three-part study, The University of Texas at Austin, www.businesswire.com/news/home/20100927005388/en/Sybase-University-Texas-Study-Reveals-Incremental-Improvement, accessed 20 May 2015.

Bekhuis, T. (2006), “Conceptual biology, hypothesis discovery, and text mining: Swanson’s legacy”, Biomedical Digital Libraries 2006 Vol. 3, No. 2, *PubMed,* 3 April, http://dx.doi.org/10.1186/1742-5581-3-2.

Bertolucci, J. (2014), “Big data hot job: Data engineer”, *Information Week,* http://www.informationweek.com/big-data/big-data-analytics/big-data-hot-job-data-engineer/d/d-id/1315579, accessed 25 May 2015.

Bloom, N. and J. Van Reenen (2007), “Measuring and explaining management practices across firms and countries”, *The Quarterly Journal of Economics*, Volume 122, No. 4, November, pp. 1351-1408, http://stanford.edu/~nbloom/MeasuringManagement.pdf.

Brave, S. (2012), “We don’t need more data scientists – just simpler ways to use big data”, Gigaom, 22 December, https://gigaom.com/2012/12/22/we-dont-need-more-data-scientists-just-simpler-ways-to-use-big-data.

Bruner, J. (2013), “Defining the industrial Internet” O'Reilly Radar, 11 January, http://radar.oreilly.com/2013/01/defining-the-industrial-internet.html.

Bruno, L.C. (2014), “Plate, punch card, and instructions for Herman Hollerith’s electric sorting and tabulating machine, ca. 1895”, *American Memory*, Library of Congress, http://lcweb2.loc.gov/cgi-bin/query/r?ammem/mcc:@field(DOCID+@lit(mcc/023)), accessed 25 May 2015.

BLS (Bureau of Labor Statistics) (2015), *Occupational Employment Statistics*, November 2014, www.bls.gov/oes/home.htm, accessed 28 June 2015.

BLS (Bureau of Labor Statistics) (2014), *Occupational Outlook Handbook, 2014-15 Edition*, US Department of Labor, www.bls.gov/ooh, accessed 21 November 2014.

Brynjolfsson, E., L.M. Hitt and H.H. Kim (2011), “Strength in numbers: How does data-driven decision making affect firm performance?”, Social Science Research Network (SSRN), 22 April, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486.

Brynjolfsson E. and L.M. Hitt (2000), “Beyond computation: Information technology, organizational transformation and business performance”, *Journal of Economic Perspectives*, Vol. 14, No. 4, pp. 23-48.

Brynjolfsson, E. and A. McAfee (2014), *The Second Machine Age – Work, Progress and Prosperity in a Time of Brilliant Technologies*, W.W. Norton & Co., New York.

Brynjolfsson, E. and A. McAfee (2012a), *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity and Irreversibly Transforming Employment and the Economy*, January, Research Brief, http://ebusiness.mit.edu/research/Briefs/Brynjolfsson_McAfee_Race_Against_the_Machine.pdf, accessed 12 May 2014.

Brynjolfsson, E. and A. McAfee (2012b), “Big data: The management revolution”, *Harvard Business Review*, October, https://hbr.org/2012/10/big-data-the-management-revolution/ar, accessed 12 May 2014.

Brynjolfsson, E and A. McAfee (2011), *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity and Irreversibly Transforming Employment and the Economy*, Digital Frontier Press, 17 October.

Carlisle, R. (2004), *Scientific American Inventions and Discoveries*, John Wiley & Sons, Inc., http://bioloskiblog.files.wordpress.com/2012/03/izumi-i-otkrica.pdf, accessed 12 May 2014.

c|net (2012), "Foxconn reportedly installing robots to replace workers", 13 November, www.cnet.com/news/foxconn-reportedly-installing-robots-to-replace-workers/.

Cingano, F. (2014), "Trends in income inequality and its impact on economic growth", *OECD Social, Employment and Migration Working Papers*, No. 163, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxrjncwxv6j-en.

Clearwater, A. and T.J. Hughes (2013), "In the beginning ... An early history of the privacy profession", *Ohio State Law Journal*, Vol. 74, p. 897, http://ssrn.com/abstract=2411814.

Coombes, A. (2013), "How to get investment advice for less online", *The Wall Street Journal,* 4 September, www.wsj.com/news/articles/SB10001424127887323477604578654101591319918.

Colvin, G. (2014), "In the future, will there be any work left for people to do?", *Fortune*, 2 June, http://fortune.com/2014/06/02/fortune-500-future/.

Cornia, G. and J. Court (2001), "Inequality, growth and poverty in the era of liberalization and globalization", UNU-WIDER, November, www.wider.unu.edu/publications/policy-briefs/en_GB/pb4/_files/78807311723331954/default/pb4.pdf.

Cowen, T. (2013), *Average is Over: Powering America Beyond the Age of the Great Stagnation*, Dutton Adult.

Cushing, E. (2013), "Amazon Mechanical Turk: The digital sweatshop", *UTNE*, January/February, www.utne.com/science-and-technology/amazon-mechanical-turk-zm0z13jfzlin.aspx, accessed 25 May 2015.

Cukier, K. and V. Mayer-Schönberger (2013), "The rise of big data – How it's changing the way we think about the world", *Foreign Affairs*, May/June.

Data Science Association (2012), "Code of conduct", www.datascienceassn.org/code-of-conduct.html, accessed 17 December 2014.

Davenport, T.H. (2014), "Cognitive technology – Replacing or augmenting knowledge workers?", CIO Journal, *The Wall Street Journal*, 18 June, http://blogs.wsj.com/cio/2014/06/18/cognitive-technology-replacing-or-augmenting-knowledge-workers/.

Davenport, T.H. and D.J. Patil (2012), "Data scientist: The sexiest job of the 21st century", *Harvard Business Review*, October, https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1, accessed 25 May 2015.

Davenport, T. (2012), "Can you live without a data scientist?", *Harvard Business Review,* https://hbr.org/2012/09/can-you-live-without-a-data-sc, accessed 25 May 2015.

Davis, A.P. et al. (2013), "A CTD–Pfizer collaboration: Manual curation of 88 000 scientific articles text mined for drug-disease and drug-phenotype interactions", *Database*, http://dx.doi.org/10.1093/database/bat080, published online 28 November, 2013.

Dewan, S. and K. Kraemer (2000), "Information technology and productivity: Evidence from country-level data", *Management Science,* Vol. 46, No. 4, pp. 548-62.

Dunne, P.E. (2014), "Mechanical aids to computation and the development of algorithms", Lecture, Computer Science Intranet, University of Liverpool, http://cgi.csc.liv.ac.uk/~ped/teachadmin/histsci/htmlform/lect4.html, accessed 25 May 2015.

*Economist Intelligence Unit* (2014), "Networked manufacturing: The digital future", *Economist Intelligence Unit* sponsored by Siemens, 7 July, www.economistinsights.com/technology-innovation/analysis/networked-manufacturing.

Economist Intelligence Unit (2012), "The deciding factor: Big data & decision making", commissioned by Capgemini, 4 June, www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/.

Elliott, S. (2014), "Anticipating a luddite revival", in *Issues in Science and Technology*, Spring, p. 27-37; see also http://issues.org/30-3/stuart/, accessed 25 May 2015.

Essinger, J. (2004), *Jacquard's Web: How a Hand-loom Led to the Birth of the Information Age*, Oxford University Press.

Farrell, C. (2013), "It's time for a negative income tax", *Bloomberg Businessweek*, 8 August, www.businessweek.com/articles/2013-08-08/its-time-for-a-negative-income-tax.

Ford, M. (2009), *The lights in the tunnel: Automation, accelerating technology and the economy of the future*, Acculant Publishing, www.thelightsinthetunnel.com, accessed 25 May 2015.

Frey, C.B. and M. Osborne (2013), *The Future of Employment: How Susceptible Are Jobs to Computerisation?*, Oxford Martin School, University of Oxford.

Friedman, M. (1962), *Price Theory: A Provisional Text*, Kessinger Publishing (facsimile reprint, 2010).

Gino, F. and B.R. Staats (2012), "Samasource: Give work, not aid", Harvard Business School NOM Unit Case No. 912-011; UNC Kenan-Flagler Research Paper No. 2013-12, 22 February 2012, http://ssrn.com/abstract=2014220, accessed 26 May 2012.

Goldin, C. and L.F. Katz (2008), *The Race Between Education and Technology*, Belknap Press.

Goos, M., A. Manning and A. Salomons (2009), "Job polarization in Europe", *American Economic Review*: Papers & Proceedings Vol. 99, No. 2, pp. 58-63, http://dx.doi.org/10.1257/aer.99.2.58.

Gurulingappa, H. et al. (2013), "Automatic detection of adverse events to predict drug label changes using text and data mining techniques", Pharmacoepidem & Drug Safety, Volume 22, Issue 11, pp. 1189-94, Wiley Online Library, http://dx.doi.org/10.1002/pds.3493.

Harris, H., S. Murphy and M. Vaisman (2013), "Analyzing the analyzers: An introspective survey of data scientists and their work", *O'Reilly Strata,* June, www.oreilly.com/data/free/files/analyzing-the-analyzers.pdf, accessed 26 May 2015.

Herlyn, E. (2012), *Einkommensverteilungsbasierte Präferenz- und Koalitionsanalysen auf der Basis selbstähnlicher Equity-Lorenzkurven - Ein Beitrag zu Quantifizierung sozialer Nachhaltigkeit*, Wiesbaden.

Herlyn, E. et al. (2015), "Big data and analytics: What are the perspectives? Reflections on the OECD project on data-driven innovation", *OECD Digital Economy Papers*, OECD Publishing, Paris, forthcoming.

Herlyn, E., and F. J. Radermacher (2014), „Was kann Marketing für die Nachhaltigkeit tun? Eine Beobachterperspektive auf die Zukunft des Sustainable Marketing", in: *Sustainable Marketing Management*: *Grundlagen und Cases*, H. Meffert, P. Kenning and M. Kirchgeorg, (eds.), Springer Gabler Verlag.

Horton, J.J. (2011), "The condition of the Turking class: Are online employers fair and honest?", *Economics Letters*, Vol. 111, pp. 10-12, http://econpapers.repec.org/article/eeeecolet/v_3a111_3ay_3a2011_3ai_3a1_3ap_3a10-12.htm, accessed 26 May 2015.

Horton, J.J. and L.B. Chilton (2010), "The labor economics of paid crowdsourcing", Proceedings of the 11th ACM Conference on Electronic Commerce, pp. 209-18.

IBM (2014), "What is a data scientist", www-01.ibm.com/software/data/infosphere/data-scientist/, accessed 25 May 2015.

IBM (2013), "IBM Watson hard at work: New breakthroughs transform quality care for patients", News release, 8 February, www-03.ibm.com/press/us/en/pressrelease/40335.wss, accessed 25 May 2015.

ILO (2009), "Updating the International Standard Classification of Occupations (ISCO)", International Labour Organization, 19 March, www.ilo.org/public/english/bureau/stat/isco/docs/d334.doc, accessed 27 May 2015.

Insight Data Engineering (2014), "Insight Data Engineering Fellows Program", white paper, http://insightdataengineering.com/Insight_Data_Engineering_White_Paper.pdf, accessed 25 May 2015.

Ipeirotis, P. (2013), "Mechanical Turk account verification: Why Amazon disables so many accounts", Blog post, 28 June, www.behind-the-enemy-lines.com/2013/06/mechanical-turk-account-verification.html.

Ipeirotis, P. (2010), "Demographics of Mechanical Turk", working paper, h www.researchgate.net/profile/Panos_Ipeirotis/publication/228140347_Demographics_of_Mechanical_Turk/links/00b7d51b0945c43fb5000000.pdf, accessed 7 December 2014.

(ISC)$^2$ (2011), *Annual Report 2010, Internet Information Systems Security Certification Consortium*, Palm Harbor, FL, www.isc2.org/uploadedFiles/(ISC)2_Public_Content/Annual_Reports/2010%20Annual%20Report.pdf.

Kan, M. (2013), "Foxconn to speed up 'robot army' deployment", *PCWorld*, 26 June, www.pcworld.com/article/2043026/foxconn-to-speed-up-robot-army-deployment-20000-robots-already-in-its-factories.html.

Kapitza, S. (2005), "Population blow-up and after: The demographic revolution and information society", Report to the Club of Rome and the Global Marshall Plan Initiative, Hamburg, www.clubofrome.org/?p=1785, accessed 26 May 2015.

Katz, D.M., M.J. Bommarito and J. Blackman (2014), "Predicting the behavior of the Supreme Court of the United States: A general approach", 21 July, http://ssrn.com/abstract=2463244 or http://dx.doi.org/10.2139/ssrn.2463244.

Kessler, A. (2004), *How We Got Here: A Silicon Valley and Wall Street Primer*, HarperBusiness, http://web.mit.edu/lin/Public/how-we-got-here.pdf, accessed 26 May 2015.

Keynes, J.M. (1930), "Economic Possibilities for our Grandchildren", W.W. Norton & Co., www.econ.yale.edu/smith/econ116a/keynes1.pdf, accessed 26 May 2015.

King, J. and R. Magoulas (2014), *2014 Data Science Salary Survey: Tools, Trends, What Pays (And What Doesn't) for Data Professionals*, O'Reilly Media, www.oreilly.com/data/free/2014-data-science-salary-survey.csp, accessed 26 May 2015.

King, J. and R. Magoulas (2013), *2013 Data Science Salary Survey: Tools, Trends, What Pays (And What Doesn't) for Data Professionals*, O'Reilly Media.

King, R.D. et al. (2004), "Functional genomic hypothesis generation and experimentation by a robot scientist", *Nature,* Vol. 427, No. 6971, pp. 247-252.

Kuonen, D. (2014), "A statistician's view on big data and data science", presentation at the Zurich Machine Learning and Data Science meetup in Zurich, Switzerland, 26 August, www.slideshare.net/kuonen/big-datadatascience-aug2014, accessed 26 May 2015.

Laney, D. (2012), "Defining and differentiating the role of the data scientist", Gartner blog, 25 March, http://blogs.gartner.com/doug-laney/defining-and-differentiating-the-role-of-the-data-scientist/.

Lee, T.B. (2015), "3D cameras are about to go mainstream: Here's why that's a big deal", Vox, 9 January, www.vox.com/2015/1/9/7520967/intel-realsense-3d-camera.

Levy, F. and R.J. Murnane (2013), "Dancing with robots: Human skills for computerized work", third way, 1 June, http://dusp.mit.edu/uis/publication/dancing-robots-human-skills-computerized-work.

Lodefalk, M. (2010), "Servicification of manufacturing - Evidence from Swedish firm and enterprise group level data", Working Papers No. 2010:3, Örebro University, School of Business, http://ideas.repec.org/p/hhs/oruesi/2010_003.html.

Lohr, S. (2009), "For today's graduate, just one word: Statistics", Inside Technology, *New York Times*, 5 August, www.nytimes.com/2009/08/06/technology/06stats.html.

Loukides, M. (2011), "What is data science? The future belongs to the companies and people that turn data into products", *O'Reilly Radar*, April, www.oreilly.com/data/free/what-is-data-science.csp.

Markoff, J. (2012), "Skilled work, without the worker", New York Times, 18 August, www.nytimes.com/2012/08/19/business/new-wave-of-adept-robots-is-changing-global-industry.html.

MGI (2011), "Big data: The next frontier for innovation, competition and productivity", McKinsey Global Institute, McKinsey & Company, June, www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

MacManu, R. (2009), "Report: Cloud-Based Email Cheapest Option for Most Companies", Read*w*rite, 6 January, www.readwriteweb.com/enterprise/2009/01/cloud-based-email-cheaper.php.

Marsan, C. (2012), "'Big data' creating big career opportunities for IT pros", Networkworld, 6 March, www.networkworld.com/article/2186578/data-center/-big-data--creating-big-career-opportunities-for-it-pros.html.

Mayer-Schönberger, V. and K. Cukier (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think,* John Murray, London.

*MIT Technology Review* (2010), "The truth about digital sweat shops", 12 January, www.technologyreview.com/view/417082/the-truth-about-digital-sweat-shops/.

Moore, G.E. (1965), "Cramming more components onto integrated circuits", *Electronics Magazine*, www.cs.utexas.edu/~fussell/courses/cs352h/papers/moore.pdf, accessed 15 May 2015.

Mozur, P. and L. Luk (2012), "Hon Hai Hits Obstacles in Push to Use Robots", *The Wall Street Journal*, 11 December, www.wsj.com/articles/SB10001424127887324024004578172022369346936.

Myers, A. (2011), "Stanford team trains computer to evaluate breast cancer", Stanford Medicine News Center, 9 November, http://med.stanford.edu/news/all-news/2011/11/stanford-team-trains-computer-to-evaluate-breast-cancer.html.

NSF (2005), *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*, National Science Foundation, www.nsf.gov/pubs/2005/nsb0540/, accessed 15 May 2015.

O'Connor, S. (2013), "Amazon unpacked", *FT Magazine*, 8 February, www.ft.com/intl/cms/s/2/ed6a985c-70bd-11e2-85d0-00144feab49a.html#slide0, accessed 24 March 2015.

OECD (2015), *Digital Economy Outlook 2015*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264232440-en.

OECD (2014a), 2014 Ministerial Council Statement, OECD Publishing, Paris, 7 May, C/MIN(2014)15/FINAL, www.oecd.org/mcm/2014-ministerial-council-statement.htm, accessed 15 May 2015.

OECD (2014b), "ICT, jobs and skills: Proposals for a research agenda", OECD Publishing, Paris, 16 June, DSTI/ICCP/IIS(2014)2.

OECD (2014c), "Cloud computing: The concept, impacts and the role of government policy", *OECD Digital Economy Papers*, No. 240, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxzf4lcc7f5-en.

OECD (2014d), "Challenges and opportunities for innovation through technology: The convergence of technologies", OECD Publishing, Paris, 25 September, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/stp(2013)15/final&doclanguage=en, accessed 26 May 2015.

OECD (2014e), "Inequality hurts economic growth, finds OECD research", OECD Newsroom, OECD Publishing, Paris, 9 December, www.oecd.org/newsroom/inequality-hurts-economic-growth.htm, accessed 15 May 2015.

OECD (2014f), *Education at a Glance 2014: OECD Indicators,* OECD Publishing, Paris, www.oecd.org/edu/eag.htm, accessed 28 June 2015.

OECD (2014g), "Skills and jobs in the Internet economy", *OECD Digital Economy Papers*, No. 242, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxvbrjm9bns-en.

OECD (2013a), "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, Paris, http://dx.doi.org/10.1787/5k47zw3fcp43-en.

OECD (2013b), *Supporting Investment in Knowledge Capital, Growth and Innovation*, OECD Publishing, Paris, www.oecd-ilibrary.org/industry-and-services/supporting-investment-in-knowledge-capital-growth-and-innovation_9789264193307-en, accessed 15 May 2015.

OECD (2013c), "Privacy Expert Group Report on the Review of the 1980 OECD Privacy Guidelines", OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3xz5zmj2mx-en.

OECD (2013d), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

OECD (2012a), *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264177338-en.

OECD (2012b), "ICT skills and employment: New competences and jobs for a greener and smarter economy", *OECD Digital Economy Papers*, No. 198, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k994f3prlr5-en.

OECD (2010), *OECD Information Technology Outlook 2010*, OECD Publishing, Paris, http://dx.doi.org/10.1787/it_outlook-2010-en.

OECD (2008), "Broadband and the economy: Ministerial background report, OECD ministerial meeting on the future of the Internet economy, Seoul", OECD Publishing, Paris, https://innovationpolicyplatform.org/document/broadband-and-economy-ministerial-background-report-oecd-ministerial-meeting-future, accessed 26 May 2015.

OECD (2005), "New perspectives on ICT Skills and Employment", OECD Publishing, Paris, www.oecd.org/internet/ieconomy/34769393.pdf, accessed 15 May 2015.

O'Reilly, T. (2014), "IoTH: The Internet of things and humans", *O'Reilly Radar*, 16 April, http://radar.oreilly.com/2014/04/ioth-the-internet-of-things-and-humans.html.

Patil, D.J. (2011), "Building data science teams", *O'Reilly Radar*, 16 September, http://radar.oreilly.com/2011/09/building-data-science-teams.html.

Piketty, T. (2014), *Capital in the Twenty-First Century*, Harvard University Press.

Provost, F. and T. Fawcett (2013), *Data Science for Business,* O'Reilly Media.

Radermacher, F.J. and B. Beyers (2007), *Welt mit Zukunft – Überleben im 21. Jahrhundert* [World with a future – Surviving in the 21st century], Murmann Verlag, Hamburg (2nd edition of "Welt mit Zukunft – Die Ökosoziale Perspektive, 2001).

Randers, J. (2012), *2052: A Global Forecast for the Next Forty Years*, Chelsea Green Publishing.

Reimsbach-Kounatze, C. (2015), "The proliferation of data and implications for official statistics and statistical agencies: A preliminary analysis", *OECD Digital Economy Papers*, No. 245, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js7t9wqzvg8-en.

Rifkin, J. (2014), *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*, Palgrave Macmillan.

Rockwell, M. (2013), "Gold in the data, but a shortage of miners", *FCW*, 17 September, http://fcw.com/Articles/2013/09/17/big-data-analyst-shortage.aspx?p=1, accessed 21 May 2015.

Rogers, S. (2012), "What is a data scientist?", *The Guardian*, Data Blog, 2 March, www.theguardian.com/news/datablog/2012/mar/02/data-scientist.

Ross, J. et al. (2010), "Who are the crowdworkers? – Shifting demographics in Amazon Mechanical Turk", *CHI EA 2010*, pp. 2863-72, www.international.ucla.edu/media/files/SocialCode-2009-01.pdf?AspxAutoDetectCookieSupport=1, accessed 15 May 2015.

Royster, S. (2013), "Working with big data", Bureau of Labor Statistics, US Department of Labor, www.bls.gov/careeroutlook/2013/fall/art01.pdf, accessed 15 May 2015.

Solte, D. (2009), *Global Financial System in Balance – Crisis as Opportunity for a Sustainable Future,* Terra-Media, Berlin.

Spiegel Online (2014), "Apple-Zulieferer: Foxconn will Arbeiter durch Roboter ersetzen", Spiegel Online, 8 July, www.spiegel.de/netzwelt/gadgets/foxconn-will-arbeiter-durch-roboter-ersetzen-a-979819.html.

Stewart-Smith, H. (2012), "Foxconn chairman compares his workforce to 'animals'", *ZDnet*, 20 January, www.zdnet.com/blog/asia/foxconn-chairman-compares-his-workforce-to-animals/776.

Taleb, N.N. (2010), *The Black Swan: The Impact of the Highly Improbable*, Random House Trade Paperbacks.

Taleb, N.N. (2005), *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*, Random House Trade Paperbacks.

Tambe, P. (2014), "Big data investment, skills, and firm value", *Management Science*, http://ssrn.com/abstract=2294077, accessed 15 May 2015.

*The Economist* (2014), "Coming to an office near you", 18 January, www.economist.com/news/leaders/21594298-effect-todays-technology-tomorrows-jobs-will-be-immenseand-no-country-ready.

*The Economist* (2012), "High-frequency trading: The fast and the furious", 25 February, www.economist.com/node/21547988.

*The Economist* (2010), "Data, data everywhere", 25 February, www.economist.com/node/15557443.

Techlist (2014), "50,000 MTurk workers who have performed techlist HITs", http://techlist.com/mturk/global-mturk-worker-map.php, accessed 15 May 2015.

TNO (2013), "Thriving and surviving in a data-driven society", 24 September, http://publications.tno.nl/publication/34610048/xcv74S/TNO-2013-R11427.pdf.

Turing, A.M. (1950), "Computing machinery and intelligence", *Mind*, pp. 433-60.

Uddin, Z. (2012), "The dystopian digital sweatshop that makes the Internet run", *AlterNet*, 11 September, www.alternet.org/labor/dystopian-digital-sweatshop-makes-internet-run.

UK National Career Service (2014), "Data analyst-statistician", Job Profiles, https://nationalcareersservice.direct.gov.uk/advice/planning/jobprofiles/Pages/dataanalyst-statistician.aspx, accessed 18 December 2014.

UN Global Pulse (2013), "Anatomy of a Pulse Lab", 19 December, www.unglobalpulse.org/anatomy-pulse-lab, accessed 17 December 2014.

Upbin, B. (2013). "IBM's Watson Gets Its First Piece of Business in Healthcare", *Forbes*, 8 February, www.forbes.com/sites/bruceupbin/2013/02/08/ibms-watson-gets-its-first-piece-of-business-in-healthcare/.

Violino, B. (2014), "The hottest jobs in IT: Training tomorrow's data scientists", *Forbes*, 26 June, www.forbes.com/sites/emc/2014/06/26/the-hottest-jobs-in-it-training-tomorrows-data-scientists/.

Voce, C. et al. (2009), "Should Your Email Live in the Cloud? An Infrastructure and Operations Analysis", Forrester, 5 January, www.forrester.com/rb/Research/should_email_live_in_cloud_infrastructure_and/q/id/42980/t/2.

Wang, S. and R.M. Summers (2012), "Machine learning and radiology", *Medical Image Analysis*, Vol. 16, No. 5, pp. 933-51, www.ncbi.nlm.nih.gov/pubmed/22465077, accessed 15 May 2015.

Wilkinson, R. and K. Pickett (2010), *The Spirit Level: Why Equality is Better for Everyone*. Penguin.

Wing, J.M. (2008), "Computational thinking and thinking about computing", Philosophical Transactions "A" of the Royal Society, www.cs.cmu.edu/afs/cs/usr/wing/www/talks/ct-and-tc-long.pdf, accessed 15 May 2015.

Wu, C.F.J. (1997), "Statistics = Data Science?", University of Michigan, http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf, accessed 15 May 2015.

Zittrain, J. (2009), "The Internet creates a new kind of sweatshop", *Newsweek*, 12 August, www.newsweek.com/internet-creates-new-kind-sweatshop-75751, accessed 15 May 2015.

## *Further reading*

OECD (2014), 2014 OECD Ministerial Statement on Climate Change, OECD Publishing, Paris, www.oecd.org/mcm/MCM-2014-Statement-Climate-Change.pdf, accessed 15 May 2015.

OECD (2013), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204256-en.

OECD (2008), "Broadband and the economy", OECD Digital Economy Papers, www.oecd-ilibrary.org/science-and-technology/broadband-and-the-economy_230450810820, accessed 28 June 2015.

*The Economist* (2012), "High-frequency trading: The fast and the furious", 25 February, www.economist.com/node/21547988, accessed 15 May 2015.

Wing, J.M. (2008), "Computational thinking and thinking about computing", Philosophical Transactions "A" of the Royal Society, www.cs.cmu.edu/afs/cs/usr/wing/www/talks/ct-and-tc-long.pdf, accessed 15 May 2015.




# Data-Driven Innovation

## Big Data for Growth and Well-Being



**Please cite this chapter as:**

OECD (2015), “Skills and employment in a data-driven economy”, in *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264229358-10-en