UPDATE OF THE EUROPEAN DATA MARKET STUDY
SMART 2016/0063

IDC Italia srl (Milan, IT)
The Lisbon Council (Brussels, BE)

Prepared for:
Katalin IMREI
Policy Officer
European Commission – DG CONNECT
Unit G1 – Data Policy & Innovation

20th June 2018
<table>
<thead>
<tr>
<th><strong>Author(s)</strong></th>
<th>Gabriella Cattaneo, Giorgio Micheletti, (IDC); David Osimo, Kasia Jakimowicz (Lisbon Council)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deliverable</strong></td>
<td>D2.3 First Interim Report</td>
</tr>
<tr>
<td><strong>Date of delivery</strong></td>
<td>20th June 2018</td>
</tr>
<tr>
<td><strong>Version</strong></td>
<td>2.0</td>
</tr>
</tbody>
</table>
| **Addressee officer** | Katalin IMREI  
Policy Officer  
European Commission, DG CONNECT  
Unit G1 — Data Policy and Innovation  
EUFO 1/178, L-2557 Luxembourg/Gasperich  
katalin.imrei@ec.europa.eu |
| **Contract ref.** | N- 30-CE-0835309/00-96                                                                         |
Contents

EXECUTIVE SUMMARY ............................................................................................................. 10

Quantifying the European Data Market – Key Facts & Figures ............................................. 10
Describing the Data Market – The Quantified Stories ............................................................ 13
Mapping the Data Market – The Data Landscape and the Data Market Monitoring Tool ........ 14
Acting Upon the Data Market – The Role of Policy ............................................................... 16

1. INTRODUCTION .................................................................................................................. 24

1.1 Objectives ......................................................................................................................... 24
1.2 Methodological Approach ............................................................................................... 25
The Indicators .......................................................................................................................... 25
The Report on Policy Conclusions ....................................................................................... 26
The Quantified Stories ........................................................................................................... 27
The EU Data Landscape Report ............................................................................................. 27
1.3 The European Data Market Monitoring Tool (updated) ................................................. 28
1.4 The Structure of this Report ......................................................................................... 28

2. QUANTIFYING THE DATA MARKET – KEY FACTS & FIGURES ..................... 30

2.1 Three future Development Paths: The Data Market at 2025 ....................................... 31
2.2 The Workforce Dimension: Data Professionals and Data Skills Gap ............................ 32
Measuring the Data Professionals ....................................................................................... 32
    Data Professionals in 2016 and 2017 .............................................................................. 32
    Data Professionals at 2025 ......................................................................................... 33
Measuring the Data Professionals Skills Gap ..................................................................... 34
2.3 The Supply- Demand Dimension: The Data Companies .............................................. 35
Measuring the Data Companies ............................................................................................ 35
    Data Companies in 2016 and 2017 ............................................................................ 35
    Data Suppliers Forecasts at 2025 .............................................................................. 37
    Data Users Forecasts at 2025 .................................................................................. 37
Measuring Data Companies’ Revenues .............................................................................. 37
    Data Companies’ Revenues in 2016 and 2017 ......................................................... 37
    Data Companies’ Revenues Forecasts at 2025 ......................................................... 38
2.4 The Business and Economic Dimension: The Data Market and the Data Economy .... 39
Measuring the Data Market ................................................................................................... 39
    The Data Market in 2016-2017 ................................................................................ 39
    The Data Market Forecasts at 2025 ......................................................................... 39
Measuring the Data Economy .............................................................................................. 40
The Data Economy in 2016 and 2017 ................................................................. 40
The Data Economy Forecasts at 2025 ............................................................... 41
2.5 The International Dimension - The Data Economy Beyond the EU – US, Brazil and Japan .............................. 42
   The U.S. ........................................................................................................ 42
   Brazil .......................................................................................................... 43
   Japan ......................................................................................................... 44
   International Overview and Comparison with the EU .................................... 44
3. DESCRIBING THE DATA MARKET – THE QUANTIFIED STORIES ....... 47
   3.1 Opening Private Data and Scientific Data for Public Interest .................. 47
   3.2 Story 2 – Opening Scientific Data for Innovation ................................. 48
   3.3 What the Quantified Stories tell us so far .............................................. 50
4. MAPPING THE DATA MARKET – DATA LANDSCAPE AND DATA MARKET MONITORING TOOL ............................................................ 51
   4.1 The EU Data Landscape ...................................................................... 51
   Major changes to the EU Data Landscape ............................................... 51
   Overview of the EU Data Landscape Database (status in January 2018): ........ 51
   Key Data Landscape companies (status in January 2018): ......................... 52
   4.2 The European Data Market Monitoring Tool ........................................ 53
5. ACTING UPON THE DATA MARKET – THE ROLE OF POLICY .......... 55
   5.1 The Role of Policy and the Future of Europe’s Data Economy: The Three Scenarios ........... 55
   Policy and the Baseline Scenario .............................................................. 56
   Policy and the High Growth Scenario ....................................................... 56
   Policy and the Challenge Scenario ........................................................... 57
   5.2 The Role of Policy in the Era of Digital Transformation ....................... 57
   5.3 Policy in the Data-Driven Society ......................................................... 58
   5.4 Europe’s International Competitiveness and the Role of Policy ................ 59
6. CONCLUSIONS .......................................................................................... 61
   6.1 Data Professionals .............................................................................. 62
   6.2 Data Companies .................................................................................. 62
   6.3 Data Market ....................................................................................... 63
   6.4 Data Economy ...................................................................................... 63
   6.5 The Data Economy beyond Europe – Brazil, Japan and US ........................ 64
   6.6 The Road Ahead – Plenty of Opportunities and a few Challenges ............ 64
METHODOLOGICAL ANNEX ........................................................................ 66
   Overview ................................................................................................... 66
Table of Figures
Figure 1: The Updated EDM Monitoring Tool.................................................................10
Figure 2: Database of Data Landscape companies (datalandscape.eu) ................................15
Figure 3: The European Data Market Monitoring Tool..................................................15
Figure 4: The European Data Market Monitoring Tool – bubble chart ..........................16
Figure 5: The Updated EDM Monitoring Tool...............................................................28
Figure 6: The Four Dimensions of the Data Market’s Key Facts & Figures .......................30
Figure 7: Data Economy by Type of Impact, EU28, 2025 (%) ...........................................42
Figure 8: Value, Growth and Incidence of the Data Economy as a Percentage of GDP in the U.S., Brazil, Japan and EU, 2016-2017 (€, Million, %) ...........................................................................45

Table of Figures
Figure 1: The Updated EDM Monitoring Tool.................................................................10
Figure 2: Database of Data Landscape companies (datalandscape.eu) ................................15
Figure 3: The European Data Market Monitoring Tool..................................................15
Figure 4: The European Data Market Monitoring Tool – bubble chart ..........................16
Figure 5: The Updated EDM Monitoring Tool...............................................................28
Figure 6: The Four Dimensions of the Data Market’s Key Facts & Figures .......................30
Figure 7: Data Economy by Type of Impact, EU28, 2025 (%) ...........................................42
Figure 8: Value, Growth and Incidence of the Data Economy as a Percentage of GDP in the U.S., Brazil, Japan and EU, 2016-2017 (€, Million, %) ...........................................................................45

Table of Figures
Figure 1: The Updated EDM Monitoring Tool.................................................................10
Figure 2: Database of Data Landscape companies (datalandscape.eu) ................................15
Figure 3: The European Data Market Monitoring Tool..................................................15
Figure 4: The European Data Market Monitoring Tool – bubble chart ..........................16
Figure 5: The Updated EDM Monitoring Tool...............................................................28
Figure 6: The Four Dimensions of the Data Market’s Key Facts & Figures .......................30
Figure 7: Data Economy by Type of Impact, EU28, 2025 (%) ...........................................42
Figure 8: Value, Growth and Incidence of the Data Economy as a Percentage of GDP in the U.S., Brazil, Japan and EU, 2016-2017 (€, Million, %) ...........................................................................45
Table of Tables

Table 1: Main Data Sources by Indicator ................................................................. 25
Table 2: Main Sources ......................................................................................... 26
Table 3: Data Professionals, 2016-2017-2020 Baseline and Growth Rates ............. 32
Table 4: Data Professionals Forecast 2025 - Total Number in the EU27 and EU28 and Growth Rates. Challenge, Baseline and High Growth Scenarios (Units, ‘000; %) ........................................ 33
Table 5: Data Professionals Skills Gap in the EU, 2016-2017-2020 Baseline and 2025 - Three Scenarios ........................................................................................................ 34
Table 6: Data Companies, 2016-2017-2020 Baseline and Growth Rates ................ 35
Table 7: Data Suppliers Forecast 2025 by Member State - Three Scenarios (Units; ‘000); CAGR 2025-2020 (%) ........................................................................................................ 37
Table 8: Data Users Forecast 2025 by Member State - Three Scenarios (Units; ‘000); CAGR 2025-2020 (%) ........................................................................................................ 37
Table 9: Data Companies’ Revenues and Growth, 2016-2017-2020 Baseline (€, Million; %) ................................................................. 38
Table 10: Data Market Value and Growth, 2016-2017-2020 Baseline (€, Million; %) ........ 39
Table 11: Data Market Forecast 2025 by Member State - Three Scenarios (€, Million; %) .................. 39
Table 12: Data Economy Value and Growth, 2016-2017-2020 Baseline and Impacts on GDP 2016-2017 (€, Million; %) ................................................................. 40
Table 13: Data Economy Forecast in 2025 and Impacts on GDP according to the Three Scenarios (€, Million; %) ......................................................................................... 41
Table 14: USA Indicators - Overview 2016, 2017 ....................................................... 42
Table 15: Brazil Indicators - Overview 2016, 2017 ....................................................... 43
Table 16: Japan Indicators - Overview 2016, 2017 ....................................................... 44
Table 17: Identification of main Factors driving the Scenarios ..................................... 71
Table 18: Indicator 1 – Data Professionals .................................................................. 72
Table 19: ISCO-08 Structure and Data Professionals .................................................. 73
Table 20: Indicator 2 – Number of Data Companies .................................................... 74
Table 21: Indicator 3 – Revenues of Data Companies .................................................. 75
Table 22: Indicator 4 – Size of the Data Market .......................................................... 76
Table 23: Indicator 5 – Value of the Data Economy ..................................................... 78
EXECUTIVE SUMMARY

This is the First Interim Study Report (Deliverable D2.3) of the Update of the European Data Market Study (SMART 2016/0063), which was entrusted in 2016 to IDC and the Lisbon Council. This report brings together the research results and the activities carried out by the contractors under:

- The First Report on Facts & Figures (D2.1) presenting an updated measurement of the European Data Market Monitoring Tool for the years 2016-2017 and forecasts to the year 2025 under three alternative scenarios;
- The First Report on Policy Conclusions (D2.2) analysing the role of policies in shaping the sizes and trends of the European Data Market and Data Economy as measured by the European Data Market Monitoring Tool.
- The key messages obtained from the quantified stories (D3.1 and D3.2) produced by the study team and focusing on the opening of private and scientific data for public interest and innovation.
- The First Data Landscape Report (Review at January 2018 – D4.1) providing an overview of the EU Data Landscape and offering an up-to-date zoom into the database of data market companies in Europe.

Quantifying the European Data Market – Key Facts & Figures

An updated Monitoring Tool

The new European Data Market Monitoring Tool leverages the tool that was used to measure the Data Market and the Data Economy during the period 2013-2016. The updated European Data Market Monitoring Tool designed by IDC captures the six main areas along which the indicators were measured to obtain the latest facts and figures about Europe’s data market and data economy. They are shown in the Figure below.

*Figure 1: The Updated EDM Monitoring Tool*
The EU Data Market and Data Economy in 2017

The pace of growth of the European Data Economy accelerated in 2017 compared to the previous years, responding to the more favourable economic climate and the pace of innovation. The number of data suppliers and data users increased, their demand of data services and products drove up the value of the Data Market, more data professionals were employed and the economic impacts resulting from this activity – the Data Economy - reached 335.5 billion Euro with an increase of 11.8% compared to the previous year, corresponding to 2.4% of the EU28 GDP.

This snapshot of the key indicators reflects a virtuous circle between demand and supply of data-driven innovation across Europe. This is demonstrated by the buoyant growth rate of the EU28 Data Market value in 2017, over 9% year-on-year, surpassing 65 billion Euro. This is a constant and significant progression from the 47 billion Euro measured for EU28 in 2013. This positive dynamic is common to all Member States, even though the pace of growth varies.

As demand grows, so does supply. The EU28 data industry measured here includes companies covering the full value stack of data, including analytics, vertical applications, cross infrastructures, ICT enablers, Data Marketplaces and industry dataspaces. Data suppliers are drawn from the ICT and Professional services industries and their number in the EU28 grew to 276 thousand in 2017, corresponding to 15% of all the companies in these 2 sectors. Their revenues increased by almost 11% in 2017 over the previous year, reaching 68.5 billion Euro, showing their good health. It should be mentioned that the U.K. alone accounted for an additional 134,000 data suppliers in 2017, given its traditional dominance of the ICT industry in Europe. The same applies to Data users (i.e. companies and organizations making an intensive use of data for business, operational and strategic decisions), which are projected to grow at 2.1% in 2017, amounting to 690,650 units in the EU28. When compared to the measurements carried out by the European Data Market Monitoring Tool over the period 2013-2015, these latest estimates show a more dynamic pictures of data companies in the EU, with growth rates constantly increasing over the past four years.

A key benefit of innovative markets is jobs creation. To design and implement data-driven innovation, enterprises are employing a wide range of skills combining mathematics and analytics, business intelligence and ICT skills, leveraging multi-disciplinary work teams. The EDM Monitoring Tool identifies these workers as data professionals, whose primary activity is to deal with data and make decisions with data. The number of data professionals in the EU28 reached 6.7 million in 2017, corresponding to 3.2% of the total workforce, with an increase of 8% over the previous year. However, the increasing demand for innovative data skills is not fully satisfied. According to our estimate of the demand-supply balance (which includes new graduates, entries from other careers and upskilling), in 2017 there was an excess demand of 449,000 data professionals, corresponding to 6% of the total skills demand in the EU28. A vacancies ratio around 5% of demand or less is considered manageable. From this point of view, this skills gap is not a very large one; what is worrying is the persistence of the data skills gap in the last years, the mismatches between the type of skills required and those offered (for example the insufficient availability of data scientists with highly sophisticated skills), and the trend towards an increase of the gap in the next years.

The EU Data Market and Data Economy in 2025

The Update of the European Data Market Study also produced key facts & figures at for the year 2025 according to three alternative evolution paths of the European Data Market and Economy, driven by different macroeconomic and framework conditions, highlighting the critical turning points to be faced in the next years by governments, businesses and social actors. The scenarios focused on

[theLisborcouncil]

[think tank for the 21st century]

[15]

[ANALYZE THE FUTURE]

[IDC]
the year 2025, taking as a reference starting point the 2020 scenarios presented in February 2017. They are built around the intersection of two main focal issues:

- **The Data Market’s pace of growth**: how fast will data-driven innovation grow in Europe? The scenarios outline the 3 alternative possibilities of slow, medium or fast pace of innovation.
- **The potential evolution of the model of data governance**, in terms of how the ownership, access, control and exploitation of data assets will be managed. To put it more bluntly: who will have power on data and what will governments do about it? The scenarios outline future models ranging between two potential extremes: on the one hand, a data governance model where a few data holders (private or public) control most of data assets; on the other hand, an open and participatory data governance model, based on sharing and transparency.

The Data Economy scenarios are positioned at the intersection of these two main focal issues and the take the following shape:

- The **Baseline scenario** is characterised by a healthy growth of data innovation, a moderate concentration of power by dominant data owners with a data governance model protecting personal data rights, and an uneven but rather wide distribution of data innovation benefits in the society;
- The **High Growth scenario** (Data-driven reality) is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society;
- The **Challenge scenario** (Digital Maze) is characterised by a low level of data innovation, a moderate level of data power concentration due to digital markets fragmentation, and an uneven distribution of data innovation benefits in the society.

In the **Baseline scenario**, the EU GDP cumulative growth average in the period 2017-2025 (+1.3%) will be 50% higher than in the Challenge scenario and 70% lower than in the High Growth scenario. This will sustain the investments in the digital economy and consumer willingness to spend. As a result, the EU27 Data Market is forecast to reach 85 billion Euro with a cumulative growth rate of 7% between 2025 and 2020. The Data Economy will reach a value of 544 billion Euro, doubling its incidence on EU GDP to 4%, compared to 2.2% in 2017. Enterprises will add more than 3 million data professionals’ positions between 2020 and 2025. However, this will increase the potential data professionals skills gap to approximately 1 million unfilled positions, corresponding to 11% of total data skills demand. The lack of skills may become a bottleneck for some enterprises or regions, as data talent “wars” will likely develop for the most skilled professionals.

In the **High Growth scenario**, the EU GDP cumulative growth average in the period 2020-2025 (+2.2%) will be 3 times higher than in the Challenge scenario and almost 2 times higher than in the Baseline scenario. This will accelerate the investments in the digital economy and consumer willingness to spend. In the European Union public and private investments will accelerate in Artificial Intelligence, advanced robotics, automation as well as new skills. As a result, the EU27 Data Market is forecast to reach 109 billion Euro with a cumulative growth rate of 7% between 2025 and 2025

---

2020. The Data Economy will reach a value of 770 billion Euro, with an incidence on EU GDP of 5.4%, compared to 2.2% in 2017. Enterprises will add more than 4 million data professionals’ positions between 2020 and 2025 (compared to 3 million in the previous scenario). However, the potential data professionals skills gap will grow exponentially to 2 million unfilled positions, corresponding to 24% of total data skills demand. This risk will need to be managed in advance, otherwise the lack of skills will become a serious constraint for data-driven companies and data suppliers.

The **Challenge scenario** will be characterised by an unexpected resistance to digital transformation by enterprises and the population due to reluctance to change, and difficulty in adopting new processes and ways of working, compounded by a relatively old working population and lack of specialist skills. As a result, the EU27 Data Market is forecast to approach the 74 billion Euro marking a Compound Annual Growth Rate of 4% between 2020 and 2025. In the same context, the Data Economy will reach a value of 377 billion Euro with an incidence on EU27 GDP of 2.9%, compared to 2.2% in 2017 or to 4% in the Baseline scenario. The number of data professionals will still increase to 8.6 million in 2025, adding 2 million data professionals’ positions compared to 2020 (a third less than in the Baseline scenario). We estimate a potential data skills gap of approximately 900 thousand unfilled positions in 2025, corresponding to 10% of total demand, as demand will still grow faster than supply. The lower supply will be due mainly to lower market entries from other careers and less upskilling-retraining initiatives, because of the lower attractiveness of the Data Market. The uneven diffusion of data innovation will result in a mismatch between demand and supply by geographical area across the Union, with unemployment in some regions and unsatisfied demand in others.

**The EU Data Market and the International Indicators**

The latest estimates of the European Data Market Monitoring Tool confirm the picture drawn by the previous measurements of the international indicators - the EU28 is the second-largest and the second most dynamic Data Economy worldwide after the U.S. and well ahead of Brazil and Japan (the other two EU international partners considered by the European Data Market Study).

Interestingly, though, Europe appears to be catching up on its gap with the U.S. and shows a renovated dynamism in some of the most significant Data Economy’s areas. In terms of data suppliers, for example, the EU can exhibit a year-on-year growth 2017-2016 of 9.2% - more than twice than in the U.S. and three times stronger than in Japan over the same period. While close to double-digit growth, the value of the Data Market in the EU28 has marked a relative halt with respect to the U.S. in the period 2016-2017 but has continued to largely outpace Brazil and, to a much lesser extent, Japan.

Where Europe shows an undoubted vitality is in the Data Economy. According to our latest estimates, direct and backward indirect impacts have grown at a much higher speed in the EU28 than in the U.S, and in the other international partners of the EU. Even more interestingly, the Data Economy as a share of GDP has marked a significant step forward in the EU28, passing from 0.42% to 0.52% of GDP (direct and backward indirect impacts only), thus growing at a rate of more than 23% year-on-year versus a U.S. growth of “only” 4.2% in the same period.

The four infographics at the end of this Executive Summary provide a comprehensive overview of the main facts and figures stemming from the Update of the European Data Market Study so far.

**Describing the Data Market – The Quantified Stories**

The quantified stories were the result of a mixed effort entailing both secondary and primary research activities and were aimed to add quantitative and qualitative evidence to the indicators
measured through the European Data Market Monitoring Tool. The stories focused on the use of data produced by the private sector and by scientific research to address public issues.

While examples of public sector data being intensively used by the private organisations abound, the opposite is less frequent, but it is real. Indeed, the free flow of data can generate maximum societal and economic benefits and there are plenty of projects and initiatives to promote data sharing both in the private sector and in research, although they are far from becoming “mainstream”. What is still unclear is how this free flow of data can be achieved at systemic level. Most of all, there is a clear need to understand the structure of objectives and incentives for different stakeholders. Pursuing the public good is not a sufficient driver for data sharing. Companies are profit driven. They share data typically by selling integrated analytics services, for instance Vodafone offers packaged services to government based on the mobility data gathered by their antennas. Companies also can provide different levels of access under freemium business models: for instance, aggregated data for free and granular data for a premium, as in the case of BBVA Data & Analytics – a data supplier company established by the Spanish bank BBVA, which offers data services to specific private industries (such as retail and tourism) against a fee but also adopts open data access models where interest third parties can get hold of a (limited) amount of data for free through a set of open APIs. Similarly, researchers are driven by the recognition of their peers. Unfortunately, at this stage data sharing does not provide direct benefits in terms of reputation – for instance, in terms of impact factor or citation rates, which could improve the career opportunities. On the other hand, companies involved in collaborative research projects are reluctant to allow data sharing, in order to maintain competitive advantage.

Data held by private companies and scientists can be invaluable for addressing societal issues, or for generating new products and services. But they are not a low hanging fruit: they require substantial investment, in some cases with the direct involvement of those who gathered the data in the first place (such as mobile operators). And any mandatory data sharing measure needs to be carefully designed in order not to hurt the emerging Data Economy: each ecosystem is building its own set of business models and organisational arrangements to fit their peculiar system of incentives, from PR to new revenues stream for companies, from increased scientific productivity to career opportunities for scientists. What is clear is that data sharing and reuse are necessary, but there is no “silver bullet”, no easy single solution to ensure it. It is a matter of designing the right policy mix of raising awareness about the different modalities, building skills, and removing existing barriers.

Mapping the Data Market – The Data Landscape and the Data Market Monitoring Tool

The first new EU Data Landscape review under the Update of the European Data Market Study was performed in January 2018 and is expected to be further updated in January 2019. At the time of the review, data landscape database included a total of 1256 companies and covered 36 countries (EU-28, Belarus, Bosnia and Herzegovina, Iceland, Moldova, Norway, Serbia, Switzerland and Ukraine). By the end of May 2018, the situation did not substantially change, with 56 new companies added to the database between February and May 2018.

As a result, among 1256 companies, 94 have been now identified as key data landscape companies in line with a set of criteria adopted.
Figure 2: Database of Data Landscape companies (datalandscape.eu)

The Update of the European Data Market Study also provided for a new improved version of the European Data Market Monitoring Tool, which, together with the presentation of the new forecasts, was launched on the datalandscape.eu around 20 April 2018. The scope of data was extended from 2013 – 2020 to 2013 – 2025 forecasts. The improved version of European Data Market Monitoring Tool presents the new approach to visualisation and a new bubble chart was introduced, allowing for personalisation and modification of graph.

Figure 3: The European Data Market Monitoring Tool

Source: http://datalandscape.eu/companies
Acting Upon the Data Market – The Role of Policy

The new technology wave surrounding the ongoing process of digital transformation is based on a common element: extracting the value of data. Data is the new commodity underlying every transaction and fuelling every new insight. The McKinsey Global Institute estimates that rising global data flows have boosted world GDP by more than 10%. But McKinsey also estimates that Europe has captured so far only 12 percent of its potential from digital technologies and is lagging behind the United States - even though the European market has the potential to be the largest digital market in the world in size and value, if investments and policy decisions will provide momentum - acting to capture the digital opportunity is therefore imperative.

Europe’s Data Market and Data Economy Evolution: Policy and the Three Scenarios

Economic growth in 2017 has bounced back in all the Member States at the fastest pace since the 2008 crisis and main public sources forecast a positive 2018. There are plenty of challenges and tensions, from Brexit to youth unemployment, but the general mood is of cautious confidence in Europe’s ability to deal with our many problems. This creates an opportunity for European businesses and political leaders to take the actions needed to sustain the growth momentum and invest in the future. And there is no field where the challenges and the potential rewards of success are higher than for digital innovation.

---

The way Europe’s Data Market and Data Economy may evolve after 2017-2018 remains open: The 2025 scenarios developed for Update of the European Data Market Study (SMART 2016/0063) present potential evolution paths of the European Data Market and Economy and are built around the intersection of two main focal issues:

- The Data Market’s pace of growth;
- The potential evolution of the model of data governance.

In the Baseline scenario policy plays a mitigated role with mixed success. Today the European Data Market is fragmented, and the diffusion of data-driven innovation is uneven, but we are on the verge of introduction of several policy measures aiming at building the Digital Single Market, first of all the General Data Protection Regulation (GDPR) and the Free Flow of non-personal Data initiative (FFoD). A more relevant policy role is foreseen under the High Growth scenario. The completion of the Digital Single Market within the forecast period is a key success factor: this requires rapid and successful implementation of the GDPR and the Free-Flow of non-personal data initiative, with true liberalisation of data flows across Europe. R&D investment at EU and national level must be particularly effective and the Network of Excellence Centres must succeed in becoming innovation beacons in every region as well as increasing the supply of scarce data skills. In short, in this scenario Europe captures in full the digital opportunity, as advised by McKinsey. The Challenge scenario is driven as much by the failure of the Digital Single Market and of innovation investments than by global economic trends. A missed approval or weak implementation of the Free Flow of non-personal Data would result in the persistence of digital barriers. The lack of increasing investments in AI, robotics and automation would create a disadvantage for the European industry.

**European Companies, Digital Transformation and the Role of Policy**

European companies find themselves in a critical stage amid this unrelenting growth of the digital transformation process with several policy challenges lying ahead. The war on talent, for example, is heating up. Europe has long faced a skills gap for ICT skills, but the digital transformation process is also adding the need of new digital skills to operational and business staff. Increasingly, these employees will be expected to have greater analytics and AI competences in order to be able to use and make sense of large data sets to improve and automate process tasks. As it will take time for the European education systems to adapt to the growing demand for these combination of skills, policy intervention will be required and organisations will need to consider a combination of internal training programs for existing staff and close cooperation with educational institutions to ensure graduates will have the required skill sets. Similarly, Europe’s evolving regulatory environment around data privacy, protection, and security will have to be closely watched to reduce possible effects and delay on the impact and timing of data monetization compared with other parts of the world. Indeed, some European industries are highly regulated – this is the case of Manufacturing and Healthcare, for instance. This, combined with a fragmented, country-based IT landscape, could inhibit the level of data access, security and governance necessary to drive a successful cognitive/AI project. Again, appropriate policy adoption and enforcement is needed to go beyond Member-State specific use cases and leverage the value of existing data across multi-company ecosystems.

**Europe’s International Competitiveness and the Role of Policy**

The international indicators draw a mixed picture: the EU still lags behind the U.S. in terms of size of the Data Market and the Data Economy but has by no means reached its potential from digital technologies. Indeed, Europe’s competitive position against the U.S. will not improve until all the measures in the Digital Single Market Strategy are implemented across all EU countries. Self-driving cars, for example, are starting to roam American streets under test conditions and the U.S.
The government has clear policy recommendations on this area. The EU has a less coordinated approach, which brings to the fore the importance of the DSM and the need to remove the barriers that still impede the free flow of data within the EU both at cross-border and at intra-company level. European companies, though, should also be adequately supported and incentivised to open up and share their data. Indeed, data users and data suppliers should be put in a position to scale within and beyond individual Member States and realise the potential of an open and effective Digital Single Market.

A significant increase in digital infrastructure investments and digital skills should also be carefully considered to deepen and expand Europe’s digital ecosystem. In this respect, particular attention should be devoted to Europe’s already vibrant next generation start-ups ecosystem, but additional efforts should be undertaken to embrace innovative digital technologies such as AI, machine-to-machine, robotics and others – the real drivers of the digital economies in the upcoming years.

Furthermore, education and skills should be step up. Worker displacement and transition would be inevitable as digitisation progresses and policy makers and business leaders alike should pay more attention and investments into Science, Technology, Engineering and Math (STEM) skills through school systems, while favouring effective retraining for mid-career workers as these transition to the future-of-work-like positions.
# The European Data Market Monitoring Tool – Key Numbers 2017 for EU28

## Indicator 1: Data Professionals

<table>
<thead>
<tr>
<th>Year</th>
<th>Million</th>
<th>Growth '17/'16</th>
<th>Share of data professionals on total EU employment, '17</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>6.7</td>
<td>8%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

## Indicator 6: Data Professionals Skills Gap

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>Growth '17/'16</th>
<th>Share of total skills demand, '17</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>449</td>
<td>5%</td>
<td>6%</td>
</tr>
</tbody>
</table>

The indicator captures the potential gap between demand and supply of data skills in Europe.

## Indicator 2: Data Companies

### Data suppliers

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>Growth '17/'16</th>
<th>Share of total companies in ICT and Professional Services, '17</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>261</td>
<td></td>
<td>15%</td>
</tr>
<tr>
<td>2017</td>
<td>276</td>
<td>5.7%</td>
<td></td>
</tr>
</tbody>
</table>

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

### Data users

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>Growth '17/'16</th>
<th>Share of total companies, '17</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>676</td>
<td></td>
<td>6.6%</td>
</tr>
<tr>
<td>2017</td>
<td>690</td>
<td>2.1%</td>
<td></td>
</tr>
</tbody>
</table>

Data users are organisations that generate, exploit, collect and analyse digital data intensively and use what they learn to improve their business.

## Indicator 3: Data suppliers’ revenues

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>Growth '17/'16</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>61.7</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>68.5</td>
<td>11%</td>
</tr>
</tbody>
</table>

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

## Indicator 4: Value of the Data Market

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>Growth '17/'16</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>59.4</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>65</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

The marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data.

## Indicator 5: Value of the Data Economy

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>Growth '17/'16</th>
<th>Share of EU GDP, '17</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>300</td>
<td></td>
<td>2.4%</td>
</tr>
<tr>
<td>2017</td>
<td>335.5</td>
<td>11.8%</td>
<td></td>
</tr>
</tbody>
</table>

The Data Economy measures the overall impacts of the data market on the economy as a whole.

*Source: EDM Monitoring Tool, IDC 2018*
## The European Data Market Monitoring Tool – Baseline Scenario 2025 for EU27

### Indicator 1: Data Professionals
- **2020:** 6.6 Million
- **2025:** 9.4 Million
- **CAGR ‘25/’20:** 7.2%

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

### Indicator 6: Data Professionals Skills Gap
- **2020:** 603 Thousand
- **2025:** 1,007 Thousand
- **CAGR ‘25/’20:** 11%
- **Share of total skills demand, ’25:** 11%

The Indicator captures the potential gap between demand and supply of data skills in Europe.

### Indicator 2: Data Companies
#### Data suppliers
- **2020:** 157 Thousand
- **2025:** 182 Thousand
- **CAGR ‘25/’20:** 3%
- **Companies Share as a % of total companies in ICT and Professional Services, ’25:** 14.5%

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

#### Data users
- **2020:** 540 Thousand
- **2025:** 585 Thousand
- **CAGR ‘25/’20:** 1.6%
- **Companies Share as a % of total companies, ’25:** 6.6%

Data users are organisations that generate, exploit, collect and analyse digital data intensively and use what they learn to improve their business.

### Indicator 3: Data Suppliers’ revenues
- **2020:** € 70 Bn
- **2025:** € 97 Bn
- **CAGR ‘25/’20:** 6.9%

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

### Indicator 4: Value of the Data Market
- **2020:** € 60 Bn
- **2025:** € 85 Bn
- **CAGR ‘25/’20:** 7.1%

The marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data.

### Indicator 5: Value of the Data Economy
- **2020:** € 366 Bn
- **2025:** € 544 Bn
- **CAGR ‘25/’20:** 8.3%
- **Share of EU GDP, ’25:** 4%

The Data Economy measures the overall impacts of the data market on the economy as a whole.

*Source: EDM Monitoring Tool, IDC 2018*
The European Data Market Monitoring Tool – High Growth Scenario 2025 for EU27

Indicator 1: Data Professionals

2020: 6.6 Million
2025: 11 Million
CAGR 2025/20: 10%

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

Indicator 6: Data Professionals Skills Gap

2020: 603 Thousand
2025: 2,116 Thousand
CAGR 2025/20: 24%
Share of total skills demand, ‘25: 19%

The indicator captures the potential gap between demand and supply of data skills in Europe.

Indicator 2: Data Companies

Data suppliers
2020: 157 Thousand
2025: 203 Thousand
CAGR 2025/20: 5%
Companies Share as a % of total companies in ICT and Professional Services, ‘25: 16%

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

Data users
2020: 540 Thousand
2025: 629 Thousand
CAGR 2025/20: 3%
Companies Share as a % of total companies, ‘25: 7%

Data users are organisations that generate, exploit, collect and analyse digital data intensively and use what they learn to improve their business.

Indicator 3: Data Suppliers’ revenues

2020: €70 Bn
2025: €133 Bn
CAGR 2025/20: 14%

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

Indicator 4: Value of the Data Market

2020: €60 Bn
2025: €109 Bn
CAGR 2025/20: 13%

The marketplace where digital data is exchanged as "products" or "services" as a result of the elaboration of raw data.

Indicator 5: Value of the Data Economy

2020: €366 Bn
2025: €769 Bn
CAGR 2025/20: 16%
Share of EU GDP, ‘25: 5%

The Data Economy measures the overall impacts of the data market on the economy as a whole.

Source: EDM Monitoring Tool, IDC 2018
### Indicator 1: Data Professionals

<table>
<thead>
<tr>
<th>Year</th>
<th>Million</th>
<th>CAGR '25/'20</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>2025</td>
<td>8.7</td>
<td>6%</td>
</tr>
</tbody>
</table>

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

### Indicator 6: Data Professionals Skills Gap

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>CAGR '25/'20</th>
<th>Share of total skills demand, '25</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>603</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2025</td>
<td>904</td>
<td>12%</td>
<td>10%</td>
</tr>
</tbody>
</table>

The indicator captures the potential gap between demand and supply of data skills in Europe.

### Indicator 2: Data Companies

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>CAGR '25/'20</th>
<th>Companies Share as a % of total companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>157</td>
<td></td>
<td>2%</td>
</tr>
<tr>
<td>2025</td>
<td>170</td>
<td>2%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Data users are organisations that generate, exploit collect and analyse digital data intensively and use what they learn to improve their business.

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousand</th>
<th>CAGR '25/'20</th>
<th>Companies Share as a % of total companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>540</td>
<td>0.8%</td>
<td>6%</td>
</tr>
<tr>
<td>2025</td>
<td>562</td>
<td></td>
<td>6%</td>
</tr>
</tbody>
</table>

### Indicator 3: Data Suppliers’ revenues

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>CAGR '25/'20</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>2025</td>
<td>77</td>
<td>2%</td>
</tr>
</tbody>
</table>

### Indicator 4: Value of the Data Market

The marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data.

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>CAGR '25/'20</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>74</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Indicator 5: Value of the Data Economy

The Data Economy measures the overall impacts of the data market on the economy as a whole.

<table>
<thead>
<tr>
<th>Year</th>
<th>€ Bn</th>
<th>CAGR '25/'20</th>
<th>Share of EU GDP, '25</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>366</td>
<td>0.6%</td>
<td>3%</td>
</tr>
<tr>
<td>2025</td>
<td>376</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: EDM Monitoring Tool, IDC 2018
### Number of Data Suppliers

<table>
<thead>
<tr>
<th>Country</th>
<th>2016</th>
<th>2017</th>
<th>Growth '17/'16</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>289,556</td>
<td>302,810</td>
<td>4.58%</td>
</tr>
<tr>
<td>Brazil</td>
<td>35,979</td>
<td>36,387</td>
<td>1.13%</td>
</tr>
<tr>
<td>Japan</td>
<td>101,612</td>
<td>104,664</td>
<td>3%</td>
</tr>
<tr>
<td>EU</td>
<td>261,450</td>
<td>276,450</td>
<td>9.2%</td>
</tr>
</tbody>
</table>

### Value of the Data Economy

#### Direct Impacts

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>Growth '17/'16</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>108,521</td>
<td>113,677</td>
<td>4.75%</td>
</tr>
<tr>
<td>Brazil</td>
<td>7,270</td>
<td>7,766</td>
<td>6.82%</td>
</tr>
<tr>
<td>Japan</td>
<td>6,157</td>
<td>6,395</td>
<td>3.86%</td>
</tr>
<tr>
<td>EU</td>
<td>290</td>
<td>298</td>
<td>2.72%</td>
</tr>
</tbody>
</table>

#### Share of GDP

<table>
<thead>
<tr>
<th>Year</th>
<th>USA</th>
<th>Brazil</th>
<th>Japan</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>0.78%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.12%</td>
</tr>
<tr>
<td>2017</td>
<td>0.81%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.12%</td>
</tr>
</tbody>
</table>

#### Backward Indirect Direct Impacts

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>Growth '17/'16</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>27,944</td>
<td>29,949</td>
<td>9.33%</td>
</tr>
<tr>
<td>Brazil</td>
<td>1,189</td>
<td>1,269</td>
<td>6.66%</td>
</tr>
<tr>
<td>Japan</td>
<td>53,509</td>
<td>65,038</td>
<td>15.55%</td>
</tr>
<tr>
<td>EU</td>
<td>2,780</td>
<td>3,303</td>
<td>19.81%</td>
</tr>
</tbody>
</table>

#### Share of GDP

<table>
<thead>
<tr>
<th>Year</th>
<th>USA</th>
<th>Brazil</th>
<th>Japan</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>0.42%</td>
<td>0.16%</td>
<td>1.2%</td>
<td>0.85%</td>
</tr>
<tr>
<td>2017</td>
<td>0.42%</td>
<td>0.16%</td>
<td>1.2%</td>
<td>0.85%</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The European Data Market Study (SMART 2013/0063) was launched by the European Commission in 2013 to measure the progress, size and trends of the European Data Economy with the objective of supporting the Data Value Chain policy of the European Commission. The study designed, developed and implemented a European Data Market Monitoring Tool providing facts and figures on the size and trends of the EU Data Market and Data Economy in the form of a series of quantitative indicators. The study also covered qualitative-aspects of the European Data Economy in the form of quantified stories investigating elements of the Data Market that were not captured by the Monitoring Tool. Finally, the European Data Market Study included a data landscaping tool offering a continuously updated picture of data companies in Europe and comprehended a series of webinars to disseminate the research results.

To continue gathering reliable and fact-based evidence on the EU Data Economy and measure the progress of the data-driven economy policies within the general framework of the Digital Single Market Strategy, the European Commission commissioned an update of the European Data Market (EDM) Study. The present document constitutes the First Interim Study Report (D2.3) of the Update of the European Data Market Study (SMART 2016/0063), which was entrusted in 2016 to IDC and the Lisbon Council. This report brings together the research results and the activities carried out by the contractors under:

- The First Report on Facts & Figures (D2.1) presenting an updated measurement of the European Data Market Monitoring Tool for the years 2016-2017 and forecasts to the year 2025 under three alternative scenarios;
- The First Report on Policy Conclusions (D2.2) analysing the role of policies in shaping the sizes and trends of the European Data Market and Data Economy as measured by the European Data Market Monitoring Tool.
- The key messages obtained from the quantified stories (D3.1 and D3.2) produced by the study team and focusing on the opening of private for public interest and scientific data for and innovation purposes.
- The First Data Landscape Report (Review at January 2018 – D4.1) providing an overview of the EU Data Landscape and offering an up-to-date zoom into the database of data market companies in Europe.

1.1 Objectives

As for the previous study, the Update of the European Data Market Study (SMART 2016/0063) pursues three main objectives closely interrelated, which together allow to develop a complete and coherent picture of the European Data Market and Data Economy. They are as follows:

- Measuring the EDM indicators, providing facts and figures on all the key features of the European Data Market and Economy, regularly updated during the life of the project, building on the taxonomy and methodology approach previously developed and successfully implemented;
- Analysing relevant issues for the development of the data ecosystem, providing Data Market stories based on factual evidence, case studies and complementary data to the EDM indicators, following on the 12 stories already published by the previous study;
• Mapping and visualising the stakeholders populating the EU Data Market, building on the stakeholders’ landscape and community developed in the previous study, and leveraging the visibility achieved by the website wwwdatalandscape.eu.

1.2 Methodological Approach

The Indicators

The measurement of each of the indicators is based on a sophisticated methodology that combines data collection, models, and desk research. Some initial assumptions are built on surveys completed during March 2015, which are supported by ongoing annual surveys. The 2015 survey includes 8 Member States, and the annual surveys include 6 Member States. The initial survey targeted potential data companies in two industries (ICT and Professional services), and data users in 11 industries. The annual update surveys target all business sectors, and company sizes greater than 10 employees. The survey is balanced to represent the mix of industries and size bands for companies in the European Union. The models used to represent expected market and company behaviour take inputs from macroeconomic indicators such as GDP and GDP growth, ICT spending, and employment.

Data sources, their use, and date updated are as follows:

Table 1: Main Data Sources by Indicator

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Updated</th>
<th>Used in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurostat Business Demographic Statistics</td>
<td>Dec 2017</td>
<td>Data professionals, Data companies, Data users</td>
</tr>
<tr>
<td>Eurostat annual structural business statistics</td>
<td>Dec 2017</td>
<td>Data professionals, Data companies, Data users</td>
</tr>
<tr>
<td>Eurostat chain linked Volumes (GDP)</td>
<td>Dec 2017</td>
<td>Data Market, Data Revenues</td>
</tr>
<tr>
<td>IDC Core IT Spending guide 2H2016</td>
<td>Jul 2017</td>
<td>Data Market, Data Revenues</td>
</tr>
<tr>
<td>IDC Worldwide Black Book (standard edition)</td>
<td>Nov 2017</td>
<td>Data Market, Data professionals, Data companies, Data Users, Data Revenues</td>
</tr>
<tr>
<td>IDC European Vertical Markets survey (2017)</td>
<td>Sep 2017</td>
<td>Data Market</td>
</tr>
<tr>
<td>IMF World Economic Outlook</td>
<td>Oct 2017</td>
<td>Data Market, Data Revenues, Data Economy</td>
</tr>
<tr>
<td>Consensus Forecasts – Consensus economics</td>
<td>Nov 2017</td>
<td>Data Market, Data Revenues, Data Economy</td>
</tr>
<tr>
<td>IT Big Data and Analytics spending Guide 2H2016</td>
<td>Aug 2017</td>
<td>Data Market</td>
</tr>
<tr>
<td>ILOSTAT statistics and databases</td>
<td>Jan 2018</td>
<td>Data Professionals</td>
</tr>
</tbody>
</table>

Additional relevant sources leveraged in this report were IDC annual Small and Medium Business and Vertical Markets end user surveys and IDC End-User IT Trends and Digital Transformation: IDC European Vertical Markets Survey 2017 whose results were used to confirm and adjust estimates, when necessary, of the number of companies that were data users and data suppliers. The detailed
data companies survey from 2017 provided a solid baseline for this estimate, and the annual end-
user survey by size and vertical market identified any notable changes from 2015 and 2016. IDC’s end-user survey asks specific questions about the actual and planned adoption of Big Data and Analytics, which gives a clear indication of trends in data use and supply.

The updated numbers of data users and data supplier companies were subsequently used to determine the updated results for the data companies’ revenues and were further combined with above mentioned sources to measure the indicators for Data Professionals, Data Professionals’ Skills Gap and Citizens’ Reliance on the Data Market for the year 2016, 2017 and for the three 2025 scenarios.

The Report on Policy Conclusions

Extensive desk research and additional literature review were conducted to produce the First Report on Policy Conclusions (D2.3) accompanying the quantitative results of the First Report on Facts & Figures (D2.1). To better investigate the role of policies in shaping the current and future development of the European Data Economy, the study team leveraged a mix of IDC research and other sources. A select list of these sources is offered in Table 2 below:

<table>
<thead>
<tr>
<th>Table 2: Main Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Document</strong></td>
</tr>
<tr>
<td>Data-Driven Innovation: Big Data for Growth and Well-being</td>
</tr>
<tr>
<td>Building a European Data Economy, COM(2017)</td>
</tr>
<tr>
<td>Open source machine-learning algorithms for the prediction of optimal cancer drug therapy</td>
</tr>
<tr>
<td>Commission Staff Working Document on the Free flow of data and emerging issues of the European Data Economy</td>
</tr>
<tr>
<td>Fair, Transparent and Accountable Algorithmic Decision-making Processes</td>
</tr>
<tr>
<td>The UK Digital Strategy, Policy paper “7. Data - unlocking the power of data in the UK economy and improving public confidence in its use”</td>
</tr>
<tr>
<td>The Value of Big Data and the Internet of Things to the UK Economy</td>
</tr>
<tr>
<td>Brexit: the EU data protection package</td>
</tr>
</tbody>
</table>

The three alternative scenarios at 2025 were developed by leveraging the results of the previous study featuring the Data Market and Data Economy future development paths at 2020. Fresh evidence was obtained through the ad-hoc workshop “European Data Economy by 2025” held by IDC in collaboration with BDVA and the BDVe project on the 20th October 2017 at the BDVA’s premises in Brussels³.

³ “Where will the European Data Economy be by 2025?” BDVA and IDC, Brussel, 20th October 2017 [http://www.bdva.eu/node/840](http://www.bdva.eu/node/840)
The workshop aimed at gathering insights from the high-level group of industry and research experts, from the BDVA community, about the potential growth paths of the European Data Economy by 2025.

The workshop was structured along two interactive sessions:

- The first session consisted of four tables of discussion dedicated to four topics: supply-demand dynamics, global megatrends, policy and regulatory factors, social factors. The participants were divided in four groups; each group spent 20 minutes at each table and then moved to the next, until all participants had discussed all factors. The discussion was focused on building a list of drivers and barriers of the Data Economy to 2025 ranked by level of relevance and likelihood (low, medium, high). The selection of the main topics was based on the scenario methodology used in this study and explained in the paragraph below.

- In the second session the participants were divided in groups and were asked to imagine potential scenarios, first individually and then in a group discussion where the scenarios were clustered to identify common and differentiating trends. The second session’s outcome was a long list of scenario ideas, which highlighted many potential alternative paths not only for Data Market growth and data-driven innovation, but also for social impacts especially concerning the potential risks of abuse and lack of control of personal data. This was a particularly interesting development, because it anticipated the increasing awareness of these social issues driven also by news such as the Cambridge Analytica-Facebook scandal.

The results of the discussion were collected and elaborated by IDC and presented in D2.1 First Report on Facts and Figures. The workshop also helped to elaborate other potential alternative social models of data society expanded under D2.2 First Report on Policy Conclusions. These models will be further investigated in the next two updates of the Update European Data Market study.

The Quantified Stories

The quantified stories were the result of a mixed effort entailing both secondary and primary research activities. Extensive secondary research on available public sources, specialised press and academic literature was undertaken to obtain an actionable and up-to-date understanding of private and scientific data for public interest and innovation together with a comprehensive picture of the phenomenon in Europe and worldwide.

In parallel, primary research was conducted to collect empirical evidence and validate the information obtained through the main desk research activities. Among the organisations interviewed, the following featured a prominent role:

- Hitachi in collaboration with City of Copenhagen and Danish Capital Region (https://www.citydataexchange.com/#/home);
- Lliander (https://www.liander.nl/over-liander/innovatie/open-data/data);
- Open Targets (https://www.opentargets.org/).

The EU Data Landscape Report

The data landscape report relied on the crowdsourcing of knowledge through an open process, where stakeholders could directly suggest the companies to be included in the database. The mapping of the data landscape focuses on EU 28. However, companies from other European
countries are also depicted in the database. The mapping exercise sought to achieve a balanced and comprehensive coverage of the different geographies, different typologies of companies (SMEs, large companies, research institutions etc.) and the different data sectors.

The report presented a revised and updated version of the database completed in 2016 for the European Data Market Study (SMART 2013/0063) which was used as the basis for the revision. As a result, a validation was undertaken for each observation in the database. The categorisation of the database was also reviewed in order to 1) capture new trends, in particular in the Analytics category, such as growing category of AI-related companies and Data Visualisation category and emergency of new category of Log Analytics, and 2) align the subcategories with the European Market Study model. Every company has been classified accordingly to the new categorisation. Section 2.2 presents the new categorisation in detail.

Once the classification had been refined, the dataset was significantly extended through desk research as well as through an input received from stakeholders (including input from Big Data Value eCosystem Project (BDVe) project). In the end, among the 1256 companies, 94 have been identified as key data landscape companies in line with a set of criteria adopted.

1.3 The European Data Market Monitoring Tool (updated)

Based on a modular and flexible structure, the European Data Market Monitoring Tool for the Update of the European Data Market Study leverages the existing tool that was used to measure the Data Market and the Data Economy during the period 2013-2016. The updated European Data Market Monitoring Tool designed by IDC is shown in the Figure below and its main components are further described in the following sections.

Figure 5: The Updated EDM Monitoring Tool

1.4 The Structure of this Report

This report is built along the following sections:

- The first section – corresponding to Chapter 2 – summarises the results of the First Report on Facts & Figures (D2.1) that was delivered in March 2018 and approved by the European Commission in April 2018.
• The second section – corresponding to Chapter 3 – provides additional qualitative and quantitative aspects on the European Data Market as obtained by the quantified stories produced by the study team in 2017 and the beginning of 2018.

• The third section – corresponding to Chapter 4 – presents an updated overview of the data landscape and interactive Data Market Monitoring Tool.

• The fourth section – corresponding to Chapter 5 focuses on the policy conclusions delivered in D2.2 in April 2018 and approved in May 2018.

• The final section provides for a set of concluding remarks drawing from all the different components (and deliverables) of the Update of the European Data Market study so far.
2. QUANTIFYING THE DATA MARKET – KEY FACTS & FIGURES

The key facts & figures stemming from the first round of measurement of the Update European Data Market Study (SMART 2016/0063) were obtained through the design and implementation of a set of indicators revolving around four dimensions:

- **The Workforce and Skills dimension including:**
  - Indicator 1.1 Number of data professionals;
  - Indicator 1.2 Employment share of data professionals;
  - Indicator 1.3 Intensity share of data professionals;
  - Indicator 6: Data professionals’ skills gaps.

- **The Supply and Demand dimension incorporating:**
  - Indicator 2.1: Number of data supplier companies;
  - Indicator 2.2: Share of data supplier companies;
  - Indicator 2.3: Number of data user companies;
  - Indicator 2.4: Share of data user companies;
  - Indicator 3.1: Revenues of data companies;
  - Indicator 3.2: Share of data companies’ revenues;

- **The Business and Economy dimension comprehending:**
  - Indicator 4: Value of the Data Market;
  - Indicator 5.1: Value of the Data Economy;
  - Indicator 5.2: Incidence of the Data Economy.

- **The International context dimension including a select number of indicators for Brazil, Japan and the US measuring:**
  - Indicator 1.1: Number of data professionals;
  - Indicator 1.2: Data professionals’ employment share;
  - Indicator 2.1: Number of data supplier companies;
  - Indicator 3.1: Revenues of data companies;
  - Indicator 4: Value of the Data Market;
  - Indicator 5.1: Value of the Data Economy (only Direct and Backward Indirect impacts);
  - Indicator 5.2: Incidence of the Data Economy on GDP (only direct and backward indirect impacts).

*Figure 6: The Four Dimensions of the Data Market’s Key Facts & Figures*
Each indicator was measured at the level of the total EU28 and total EU27 (excluding the U.K.) for all 28 EU Member States, when available and applicable; industry-specific and company-size views were also offered with indicators provided by industry sector and company size bands, when possible.

2.1 Three future Development Paths: The Data Market at 2025

The key facts & figures obtained through the measurement of the above-listed indicators were produced for the years 2016 and 2017 as well as for the year 2025 according to three potential future scenarios of the European Data Market and Economy, driven by different macroeconomic and framework conditions. The scenarios at 2025 took as a reference point the 2020 scenarios presented in February 2017. While the 2020 scenarios were mainly differentiated by different demand-supply dynamics, the 2025 scenarios were driven by the intersection of two main evolution paths (focal issues):

- the high or low pace of diffusion of data-driven innovation, driven by demand-supply dynamics, and its impact on economic growth;
- the high or low concentration of power in the access, control and exploitation of data assets, that is the social model of data governance. High power concentration means that a few leading data holders control most of the data assets; the opposite is an open and participatory social model of data sharing and management. In a centralised scenario, data holders also tend to appropriate a high share of benefits from data innovation, while in a decentralised scenario the benefits are more equally shared.

This analysis highlights the critical turning points to be faced in the next years by governments, businesses and social actors in the development of the European Data Economy. The combination of alternative social and economic trends results in the following scenarios:

- The Baseline scenario is characterised by a healthy growth of data innovation, a moderate concentration of power by dominant data owners with a data governance model protecting personal data rights, and an uneven but rather wide distribution of data innovation benefits in the society;
- The High Growth scenario is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society;
- The Challenge scenario is characterised by a low level of data innovation, a moderate level of data power concentration due to digital markets fragmentation, and an uneven distribution of data innovation benefits in the society.

These scenarios underline that there are relevant choices to be made in the next years about the social and economic governance model of the Data Market, in order to maximise the chances of harnessing the power of data for economic growth together with an open, transparent and shared model of data governance and control.

---

2.2 The Workforce Dimension: Data Professionals and Data Skills Gap

Measuring the Data Professionals

Data professionals⁵ are workers who collect, store, manage, and/or analyse, interpret, and visualise data as their primary or as a relevant part of their activity. Data professionals must be proficient with the use of structured and unstructured data, should be able to work with a huge amount of data and be familiar with emerging database technologies.

Data Professionals in 2016 and 2017

All in all, data professionals were estimated at a total of 5.3 million in the EU27 and at 6.7 million in the EU28 in 2017, thus marking a significant increase in 2017 over the previous year (8.2% and 8% year-on-year respectively). When compared to the year 2017, the 2020 Baseline scenario outlined in the previous European Data Market Study (SMART 2013/0063) would register a Compound Annual Growth Rate (CAGR) of 7.9% and 7.5% at the level of EU27 and EU28 respectively. More interestingly, the employment share and the intensity share components of the data professionals’ indicator are also expected to significantly improve in 2017 and 2020 if compared to our estimates in 2016 (now estimated at 3% and 3.4% in 2007 and 2020 in the EU27 and 3.2% and 3.6% for the same years in the EU28), thus confirming the positive evolution of the workforce involved in data-related professions over the period under consideration.

Table 3: Data Professionals, 2016-2017-2020 Baseline and Growth Rates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>EU27</td>
<td>Number of data professionals</td>
<td>Total number of data professionals in EU (000s)</td>
<td>4,875</td>
<td>5,273</td>
<td>6,619</td>
<td>8.2%</td>
<td>7.9%</td>
</tr>
<tr>
<td>1.1</td>
<td>EU28</td>
<td>Number of data professionals</td>
<td>Total number of data professionals in EU (000s)</td>
<td>6,187</td>
<td>6,685</td>
<td>8,309</td>
<td>8.0%</td>
<td>7.5%</td>
</tr>
<tr>
<td>1.2</td>
<td>EU27</td>
<td>Employment share of data professionals</td>
<td>Share of data professionals on total employment in EU (%)</td>
<td>2.8%</td>
<td>3.0%</td>
<td>3.4%</td>
<td>5.3%</td>
<td>4.2%</td>
</tr>
<tr>
<td>1.2</td>
<td>EU28</td>
<td>Employment share of data professionals</td>
<td>Share of data professionals on total employment in EU (%)</td>
<td>3.1%</td>
<td>3.2%</td>
<td>3.6%</td>
<td>5.3%</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

⁵ The previous European Data Market Study (SMART 2013/0063) included an indicator measuring “Data Workers”, which was based on a similar, but slightly more restrictive definition. In this updated study we have decided to measure “Data Professionals”, that is workers with a wider range of data-related roles. Indeed, data professionals are not only data technicians, but also users who, based on sophisticated tools, take decisions about their business or activities after having analysed and interpreted available data.
Data Professionals at 2025

The number of data professionals in both the EU27 and EU28 is forecast to grow significantly under all the new three scenarios at 2025 as the use of data-driven innovation is expected to grow under unabatedly even under the less economically favourable scenario. A steady progression of the number of data professionals continued to emerge from our latest estimates. Under the new Baseline scenario, data professionals are expected to amount to 9.4 million in the EU27 and 11.5 million in the EU28 at 2025, thus representing a solid growth rate between 6.7% and 7.2% over the 2017-2025 period.

In the new Challenge and High Growth scenarios, data professionals would be more than 8.6 million in the EU27 and 10.5 million in the EU28 and 10.9 million and 13.4 million respectively. Under all new scenarios, the CAGR over the period 2017-2025 would be in line with the CAGR featured by the Data Market growth, thus confirming again the close relationship between the two variables.

Table 4: Data Professionals Forecast 2025 - Total Number in the EU27 and EU28 and Growth Rates. Challenge, Baseline and High Growth Scenarios (Units, '000; %)

<table>
<thead>
<tr>
<th>N.</th>
<th>Region</th>
<th>Name</th>
<th>Description</th>
<th>2025 Challenge</th>
<th>2025 Baseline</th>
<th>2025 High Growth</th>
<th>CAGR Challenge Scenario</th>
<th>CAGR Baseline Scenario</th>
<th>CAGR High Growth Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>EU27</td>
<td>Number of data professionals</td>
<td>Total number of data professionals in EU (000s)</td>
<td>8,641</td>
<td>9,366</td>
<td>10,916</td>
<td>5.5%</td>
<td>7.2%</td>
<td>10.5%</td>
</tr>
<tr>
<td>1.1</td>
<td>EU28</td>
<td>Number of data professionals</td>
<td>Total number of data professionals in EU (000s)</td>
<td>10,487</td>
<td>11,477</td>
<td>13,450</td>
<td>4.8%</td>
<td>6.7%</td>
<td>10.1%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018
Measuring the Data Professionals Skills Gap

The **Data Professionals Skills Gap** indicator captures the potential gap between demand and supply of data skills in Europe, since the lack of skills may become a barrier to the development of the data industry and the rapid adoption of data-driven innovation. It is based on a model balancing the main sources of data skills (from the education system and re-training and other carriers) with the estimated demand (by all data companies).

This indicator continues to signal an imbalance between demand and supply of data skills in Europe. In the year 2017, given the strong increase of demand from 2016, the estimated gap reached approximately 449,000 unfilled positions, corresponding to 6% of total demand. By 2020 we expect the gap to expand to 1 million unfilled positions in the EU27.

The data skills gap will continue to grow in all three forecast scenarios, reaching 1 million unfilled positions in 2025 in the EU27 Baseline scenario (11% of demand) but up to over 2 million in the EU27 High Growth scenario (19% of demand). In the Challenge scenario the 2025 gap is slightly lower than in the Baseline. The main trends explaining these variations are the following:

- Gradual increase of the number of graduates in the period 2020-2025, thanks to increasing awareness of the market potential and to the policies promoting STEM and data skills education. These policies were launched in several countries since 2015 and their effects are starting to be seen in the forecast period. For example, Ireland has already seen an increase of 54% of level 8 graduates in the STEM field from 2012 to 2018;
- Increase of inflows in the data skills market from other careers, upskilling and re-training initiatives, due to higher attractiveness of the ICT career in the Baseline and even more so in the High Growth scenario;
- In the Challenge scenario, lower inflows to the data skills market from other careers, upskilling and re-training initiatives, because of the slower development of the Data Market;
- In all 3 scenarios, fast growth of data professionals demand that overcomes supply growth.

| Table 5: Data Professionals Skills Gap in the EU, 2016-2017-2020 Baseline and 2025 - Three Scenarios |

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Actual</th>
<th>Baseline</th>
<th>Baseline Scenario</th>
<th>Challenge Scenario</th>
<th>High Growth Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Data professionals skills gap</td>
<td>Gap between demand and supply of data professionals, N, 000s</td>
<td>EU27</td>
<td>EU28</td>
<td>2016</td>
<td>2017</td>
<td>2020</td>
</tr>
<tr>
<td>6.1</td>
<td>Data professionals skills gap</td>
<td>Gap between demand and supply of data professionals, N, 000s</td>
<td>EU27</td>
<td>EU28</td>
<td>343</td>
<td>378</td>
<td>603</td>
</tr>
</tbody>
</table>

---

2.3 The Supply-Demand Dimension: The Data Companies

Measuring the Data Companies

Data companies are organisations that are directly involved in the production, delivery and/or usage of data in the form of digital products, services and technologies. They can be both data suppliers’ and data users’ organisations:

- **Data suppliers** have as their main activity the production and delivery of digital data-related products, services, and technologies. They represent the supply side of the Data Market.
- **Data users** are organisations that generate, exploit, collect and analyse digital data intensively and use what they learn to improve their business. They represent the demand side of the Data Market.

Data Companies in 2016 and 2017

The number of data suppliers continued to grow at a faster pace than the number of data users: the former is estimated at almost 142,000 in the EU27 and at more than 276,000 units in the EU28, thus exhibiting a year-on-year growth of 5.7% in 2017. Data users, instead, are projected to grow at 2.1% in 2017, amounting to almost 516,000 in the EU27 and to 690,650 units in the EU28. If compared to the measurements carried out by the European Data Market Monitoring Tool over the period 2013-2016 these latest estimates show a more dynamic picture of data companies in the EU, with growth rates constantly increasing over the past four years.

This positive dynamic is reflected in the percentage shares of data companies over the total number of companies in Europe. The share of data suppliers on total companies in the ICT and Professional services industries is now estimated at 11.5% in the EU27 (up 0.6 percentage points vis-à-vis the previous year) and 15% in the EU28 (up 0.8 percentage points), while the data users’ penetration rates are up a modest 0.1 percentage point in 2017 in both the EU27 and EU28.

Table 6: Data Companies, 2016-2017-2020 Baseline and Growth Rates

<table>
<thead>
<tr>
<th>Indicator 2 – Data Companies 2016-2017-2020 Baseline and Growth Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.</td>
</tr>
<tr>
<td>2.1</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Number of data suppliers</td>
<td>Total number of data suppliers measured as legal entities based in the EU (000s)</td>
<td>EU28</td>
<td>261,450</td>
<td>276,450</td>
<td>305,600</td>
<td>5.7%</td>
<td>3.4%</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Share of data suppliers</td>
<td>% share of data companies on total companies in the ICT and Professional services industries</td>
<td>EU27</td>
<td>10.9%</td>
<td>11.5%</td>
<td>12.8%</td>
<td>5.6%</td>
<td>3.5%</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Number of data users</td>
<td>Total number of data users in the EU, measured as legal entities based in one EU country</td>
<td>EU27</td>
<td>505,950</td>
<td>516,350</td>
<td>539,800</td>
<td>2.1%</td>
<td>1.5%</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Number of data users</td>
<td>Total number of data users in the EU, measured as legal entities based in one EU country</td>
<td>EU28</td>
<td>676,150</td>
<td>690,650</td>
<td>721,850</td>
<td>2.1%</td>
<td>1.5%</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Share of data users</td>
<td>% share of data users on total companies in the EU industry</td>
<td>EU27</td>
<td>5.7%</td>
<td>5.8%</td>
<td>6.1%</td>
<td>1.8%</td>
<td>1.6%</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Share of data users</td>
<td>% share of data users on total companies in the EU industry</td>
<td>EU28</td>
<td>6.5%</td>
<td>6.6%</td>
<td>6.9%</td>
<td>1.9%</td>
<td>1.6%</td>
<td></td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018
Data Suppliers Forecasts at 2025

The outlook for data suppliers is continued growth, but the baseline growth to 2025 does not fully match the growth seen in 2020 according to our previous Baseline scenario – primarily because the market is larger and so growth becomes more difficult. In spite of this, there is much higher growth for the larger data supplier companies because investment as a data supplier requires resources not as readily available to smaller companies. Larger companies can afford individuals and departments whose sole purpose is to address the Data Market, while in smaller companies the development role often falls to individuals who have other responsibilities.

Table 7: Data Suppliers Forecast 2025 by Member State - Three Scenarios (Units; ’000); CAGR 2025-2020 (%)

<table>
<thead>
<tr>
<th>Region</th>
<th>2025 Challenge</th>
<th>2025 Baseline</th>
<th>2025 High Growth</th>
<th>CAGR 2025/2020 Challenge Scenario (%)</th>
<th>CAGR 2025/2020 Baseline Scenario (%)</th>
<th>CAGR 2025/2020 High Growth Scenario (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU27</td>
<td>170,100</td>
<td>182,050</td>
<td>202,800</td>
<td>1.6%</td>
<td>3.0%</td>
<td>5.2%</td>
</tr>
<tr>
<td>EU28</td>
<td>338,700</td>
<td>357,950</td>
<td>410,900</td>
<td>2.1%</td>
<td>3.2%</td>
<td>6.1%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018

Data Users Forecasts at 2025

Long term growth in the number of data user companies is highest in the data intense industries such as Professional services and ICT, and lowest in Mining and Construction, as well as Education. The largest companies show the highest growth in adoption as the Data Economy will be crucial to their success and competitive advantage – without a data-oriented approach to business and business decisions these companies will not see the opportunities their competitors see and so not grow at the same rate. However, these larger companies are a small share of the overall number of companies so although the number will grow at a compound rate of 25% to 2025, compared with 1.7% for those in the smaller size band, they do not add significantly to the total number of data companies.

Table 8: Data Users Forecast 2025 by Member State - Three Scenarios (Units; ’000); CAGR 2025-2020 (%)

<table>
<thead>
<tr>
<th>Region</th>
<th>2025 Challenge</th>
<th>2025 Baseline</th>
<th>2025 High Growth</th>
<th>CAGR 2025/2020 Challenge Scenario (%)</th>
<th>CAGR 2025/2020 Baseline Scenario (%)</th>
<th>CAGR 2025/2020 High Growth Scenario (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU27</td>
<td>562,250</td>
<td>585,150</td>
<td>629,250</td>
<td>0.8%</td>
<td>1.6%</td>
<td>3.1%</td>
</tr>
<tr>
<td>EU28</td>
<td>758,100</td>
<td>786,650</td>
<td>853,200</td>
<td>1.0%</td>
<td>1.7%</td>
<td>3.4%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018

Measuring Data Companies’ Revenues

Data companies’ revenues correspond to the aggregated value of all the data-related products and services generated by Europe-based data suppliers, including exports outside the EU.

Data Companies’ Revenues in 2016 and 2017

Data suppliers’ revenues have increased by almost 11% in 2017 to reach 52.3 billion Euro in the EU27 and 68.5 billion Euro in the EU28 – a constant increase since 2013 according to our initial measurements and the Monitoring Tool. The share of the data companies’ revenues on the total...
companies’ revenues in the ICT and Professional services sectors has grown as a result and is now stable at 3.2% in both the EU27 and the EU28.

Table 9: Data Companies’ Revenues and Growth, 2016-2017-2020 Baseline (€, Million; %)

<table>
<thead>
<tr>
<th>Indicator 3 — Data Companies’ Revenues and Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>----</td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td>3.2</td>
</tr>
<tr>
<td>3.2</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018

Data Companies’ Revenues Forecasts at 2025

Data revenues are expected to follow the Data Market, as imports and exports of data tools and services tend to follow each other. Forecasting data companies’ revenues shows an expected annual growth rate out to 2025 of 8.3% and of 7.3% from 2020 to 2025 - easily outpacing the growth of the total ICT market over the same period (expected to be 1.6% from 2016 to 2025 Baseline). The smaller Member States show the highest long-term growth as they have a smaller base from which to grow, but the larger Member States will make the biggest overall contribution to the Data Economy out to 2025.

Data Companies’ Revenues Forecast 2025 by Member State - Three Scenarios (€, Million; %)

<table>
<thead>
<tr>
<th>Indicator 3 — Data Companies’ Revenues - Forecast 2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>2025 Challenge</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Total EU27</td>
</tr>
<tr>
<td>Total EU28</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018
2.4 The Business and Economic Dimension: The Data Market and the Data Economy

Measuring the Data Market

*The Data Market* is the marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data.

The Data Market in 2016-2017

The value of the Data Market in 2017 for both the EU27 and the EU28 is showing a buoyant growth rate of more than 9% year-on-year and is expected to surpass the threshold of 60 billion Euro in 2020 in the EU27 according to the Baseline scenario described in our previous study – a constant and significant progression if we consider that the total amount of the Data Market in the EU27 was estimated at 42.6 billion Euro in 2015 in our previous study and that our current estimates measures the Data Market at almost 46.2 billion Euro in 2016.

*Table 10: Data Market Value and Growth, 2016-2017-2020 Baseline (€, Million; %)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>EU27</td>
<td>Value of the Data Market</td>
<td>Estimate of the overall value of the Data Market</td>
<td>46,183</td>
<td>50,438</td>
<td>60,254</td>
<td>9.2%</td>
</tr>
<tr>
<td>4.1</td>
<td>EU28</td>
<td>Value of the Data Market</td>
<td>Estimate of the overall value of the Data Market</td>
<td>59,496</td>
<td>65,038</td>
<td>77,407</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018

The Data Market Forecasts at 2025

In 2025 the value of the Data Market under the new High Growth scenario is expected to largely double in size with respect to its 2017 estimates in both the EU27 and the EU28. This will correspond to a considerable CAGR for the period 2020-2025 of 12.7% and 13.6% in the EU27 and the EU28 respectively. According to our new 2025 Baseline scenario, the Data Market will amount to more than 85 billion Euro in the EU27, against 50.4 billion Euro in 2017 (a 7.1% CAGR 2020-2025), while under the Challenge scenario the Data Market will still represent 73.7 billion Euro, growing at a compound annual growth rate of 4.1% from 2020. The Data Market growth will therefore continue unabated in 2025, confirming the trend set out in 2013-2014 while elaborating our initial results of the European Data Market Study (SMART 2013&0063).

*Table 11: Data Market Forecast 2025 by Member State - Three Scenarios (€, Million; %)*

<table>
<thead>
<tr>
<th>Indicator 4 — Data Market - Forecast 2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>2025 Challenge</td>
</tr>
<tr>
<td>Total EU27</td>
</tr>
<tr>
<td>Total EU28</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018
Measuring the Data Economy

The **Data Economy** measures the overall impacts of the Data Market on the economy as a whole. It involves the generation, collection, storage, processing, distribution, analysis elaboration, delivery, and exploitation of data enabled by digital technologies.

The Data Economy **includes the direct, indirect, and induced effects of the Data Market on the economy.**

- **The direct impacts:** are the initial and immediate effects generated by the data suppliers; they represent the activity potentially engendered by all businesses active in the data production. The quantitative direct impacts will then be measured as the revenues from data products and services sold, i.e. the value of the Data Market.
- **The indirect impacts:** are the economic activities generated along the company’s supply chain by the data suppliers. There are two different types of indirect impacts: the backward indirect impacts and the forward indirect impacts.
- **The induced impacts:** include the economic activity generated in the whole economy as a secondary effect.

**The Data Economy in 2016 and 2017**

The value of the Data Economy for the EU28 is to exceed the threshold of 300 billion Euro in 2016, and the estimated 2017 growth rate of 12% highlights a faster growth, considering the 7% for 2016. The share of overall impacts on GDP is expected to grow from 2.2% in 2016 to 2.4% in 2017. Results for EU27 and EU28 are similar, but it is worth highlighting the difference in the 2020/2017 CAGR and in 2017 growth rates for the two regions. EU27 rates are higher than EU28, highlighting uncertainty related to Brexit is affecting the U.K. which is showing lower than average growth rates and 2020/2017 CAGR.

**Table 12: Data Economy Value and Growth, 2016-2017-2020 Baseline and Impacts on GDP 2016-2017 (€, Million; %)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EU27</td>
<td>Value of the Data Economy and Impacts on EU GDP</td>
<td>243,205</td>
<td>267,006</td>
<td>365,761</td>
<td>9.8%</td>
<td>11.1%</td>
<td>2.1%</td>
<td>2.2%</td>
</tr>
<tr>
<td>EU28</td>
<td>Value of the Data Economy and Impacts on EU GDP</td>
<td>305,977</td>
<td>335,618</td>
<td>452,190</td>
<td>9.7%</td>
<td>10.4%</td>
<td>2.2%</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC January 2018
The Data Economy Forecasts at 2025

The Data Economy in 2025 is expected to near 700 billion Euro under the Baseline scenario impacting more than 4.2% over the EU28 and to arrive at 544 billion Euro in the EU27 (with a share on GDP of 5.4%)

The CAGR 2020/2025 in the EU28 is higher (17.2%) than the CAGR 2017/2025 (14.6%), thus highlighting an acceleration of growth after 2020, that will make the Data Economy for EU28 surpass the threshold of EUR 1 trillion, and accounting for 6% of the GDP at 2025. Similarly, the Challenge scenario will see a slowdown of the economic effects, from 4.3% CAGR 2017/2025 to 0.8% CAGR 2020/2025, with the Data Economy being well below 500 billion Euro.

Table 13: Data Economy Forecast in 2025 and Impacts on GDP according to the Three Scenarios (€, Million; %)

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>2025 Challenge Scenario</th>
<th>2025 Baseline Scenario</th>
<th>2025 High Growth Scenario</th>
<th>Impacts on GDP 2025 Challenge Scenario</th>
<th>Impacts on GDP 2025 Baseline Scenario</th>
<th>Impacts on GDP 2025 High Growth Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Value of the Data Economy EU27</td>
<td>Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy</td>
<td>376,520</td>
<td>544,241</td>
<td>769,505</td>
<td>2.9%</td>
<td>4.0%</td>
<td>5.4%</td>
</tr>
<tr>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Value of the Data Economy EU28</td>
<td>Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy</td>
<td>470,402</td>
<td>669,197</td>
<td>1,001,073</td>
<td>3.0%</td>
<td>4.2%</td>
<td>6.0%</td>
</tr>
<tr>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC January 2018

As in the previous study, this report provides a detailed insight of the Data Economy by type of impact – direct, indirect and induced impacts. The composition of impacts changes along time, from 2017 (figure below) to 2025, in favour of induced impacts, this revealing the effects of data access, data product and services exchange, and data value distribution in the economy. Indeed, induced impacts in 2025 account for a share of 56% (both in EU27 and in EU28) from around 30% in 2017. Indirect impacts in turn will lose around 11% of share, but still in 2025 accounting for a very high percentage (34%). With respect to 2017, in which the indirect impacts are the most relevant, forward impacts in particular, in 2025 induced impacts will increase, reaching a share similar to the one of the indirect impacts.
2.5 The International Dimension - The Data Economy Beyond the EU – US, Brazil and Japan

The U.S.

The number of data professionals, as well as their share on the country’s total employment base, marks the strongest growth in 2017 vis-à-vis the other two international partners under consideration (10% and 8.7% year-on-year growth in the number of data professionals and employment share in 2017 over 2016 respectively). The same applies for the data supplier companies’ indicators, with the highest increase of data suppliers in 2017 (4.6% vs. 1.1% and 3% in Brazil and Japan respectively) and an even stronger rise in their associated revenues and in the value of the Data Market in 2017 over 2016 (12.7% vs. 4.3% and 8.7% in Brazil and Japan respectively). Accordingly, the U.S.’ overall Data Economy (direct and backward indirect impacts only) has increased by almost 7% in 2017 gaining a 4.3% year-on-year in incidence on GDP and now representing more than 0.8% of the country’s GDP.

Table 14: USA Indicators - Overview 2016, 2017

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>USA – Indicators’ Overview Metrics</th>
<th>2016</th>
<th>2017</th>
<th>Growth 2017/2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Number of Data professionals</td>
<td>Total Number of Data professionals (Thousands)</td>
<td>12,732</td>
<td>14,012</td>
<td>10.05%</td>
</tr>
<tr>
<td>1.2</td>
<td>Data professionals’ employment share</td>
<td>% of Data professionals on total employment</td>
<td>8.42%</td>
<td>9.15%</td>
<td>8.66%</td>
</tr>
<tr>
<td>2.1</td>
<td>Number of Data Suppliers</td>
<td>Total number of data supplier companies (000s)</td>
<td>289,556</td>
<td>302,810</td>
<td>4.58%</td>
</tr>
<tr>
<td>3.1</td>
<td>Revenues of Data Companies</td>
<td>Total revenues generated by companies specialized in the supply of data-related products and services (Million €)</td>
<td>129,173</td>
<td>145,546</td>
<td>12.67%</td>
</tr>
</tbody>
</table>
Brazil has three years of weak growth, with 2015 and 2016 showing GDP declines, according to the IMF GDP data. 2017 show more promise, with GDP growth forecast, but there is momentum to overcome from the three low growth years, and investment is weak. However, the growth in data professionals, data companies, and the Data Economy is inevitable because of the benefits that come from digitally transforming companies. Brazil outlook is more positive for 2017, but the country still has some catching up to do when compared with European investment in the Data Economy.

Data professionals and their associated share on total employment in Brazil have grown between 1.4% and 2.1%, the number of data supplier companies, their related revenues and the accompanying Data Market have grown of a considerable 4.3%. As a result, the incidence of the economy on Brazil’s GDP will also increase but only of a modest 0.4% year-on-year (with the Data Economy – direct and backward indirect impacts – representing just 0.16% of Brazil’s GDP in 2017).

See Table below for an overview of Brazil’s growth rates in 2017.

Source: European Data Market Monitoring Tool, IDC 2018

### Table 15: Brazil Indicators - Overview 2016, 2017

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Metrics</th>
<th>2016</th>
<th>2017</th>
<th>Growth rate 2017/2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Number of Data professionals</td>
<td>Total Number of Data professionals (Thousands)</td>
<td>1,160</td>
<td>1,176</td>
<td>1.37%</td>
</tr>
<tr>
<td>1.2</td>
<td>Data professionals’ employment share</td>
<td>% of Data professionals on total employment</td>
<td>1.81%</td>
<td>1.84%</td>
<td>2.13%</td>
</tr>
<tr>
<td>2.1</td>
<td>Number of Data Suppliers</td>
<td>Total number of data supplier companies (000s)</td>
<td>35,979</td>
<td>36,387</td>
<td>1.13%</td>
</tr>
<tr>
<td>3.1</td>
<td>Revenues of Data Companies</td>
<td>Total revenues generated by companies specialized in the supply of data-related products and services (Million €)</td>
<td>6,049</td>
<td>6,310</td>
<td>4.31%</td>
</tr>
<tr>
<td>4.1</td>
<td>Value of the Data Market</td>
<td>Estimate of the overall a value of the Data Market (Million €)</td>
<td>6,049</td>
<td>6,310</td>
<td>4.31%</td>
</tr>
<tr>
<td>4.2</td>
<td>Value of the Data Economy</td>
<td>Direct Impacts (Million €)</td>
<td>6,157</td>
<td>6,395</td>
<td>3.86%</td>
</tr>
<tr>
<td></td>
<td>(Only Direct and Backward Indirect impacts)</td>
<td>Backward Indirect Impacts (Million €)</td>
<td>290</td>
<td>298</td>
<td>2.72%</td>
</tr>
<tr>
<td>4.3</td>
<td>Incidence of the Data Economy on GDP</td>
<td>Ratio between value of the Data Economy and GDP (%)</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.38%</td>
</tr>
</tbody>
</table>
Japan

The indicators measuring the state of the Data Market and the Data Economy in Japan have all registered a significant growth in 2017 over the previous year. The number and employment share of data professionals rose by 8% and 4% respectively in 2017, while the number of data suppliers increased by 3% generating a significant growth in revenues and Data Market of 8.6%. The Data Economy has therefore also marked a positive development in Japan with an incidence on GDP now at 0.95% of the country’s GDP – a growth of 1.6% with respect to the incidence in 2016. The Table below presents the main growth rates for Japan’s Data Economy.

Table 16: Japan Indicators - Overview 2016, 2017

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Metrics</th>
<th>2016</th>
<th>2017</th>
<th>Growth rate 2017/2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Number of Data professionals</td>
<td>Total Number of Data professionals (Thousands)</td>
<td>3,740</td>
<td>4,040</td>
<td>8.03%</td>
</tr>
<tr>
<td>1.2</td>
<td>Data professionals’ employment share</td>
<td>% of Data professionals on total employment</td>
<td>5.82%</td>
<td>6.05%</td>
<td>4.00%</td>
</tr>
<tr>
<td>2.1</td>
<td>Number of Data Suppliers</td>
<td>Total number of data supplier companies (000s)</td>
<td>101,612</td>
<td>104,664</td>
<td>3.00%</td>
</tr>
<tr>
<td>3.1</td>
<td>Revenues of Data Companies</td>
<td>Total revenues generated by companies specialized in the supply of data-related products and services (Million €)</td>
<td>25,513</td>
<td>27,723</td>
<td>8.66%</td>
</tr>
<tr>
<td>4.1</td>
<td>Value of the Data Market</td>
<td>Estimate of the overall a value of the Data Market (Million €)</td>
<td>25,513</td>
<td>27,723</td>
<td>8.66%</td>
</tr>
<tr>
<td>4.2</td>
<td>Value of the Data Economy (Only Direct and Backward Indirect impacts)</td>
<td>Direct Impacts (Million €)</td>
<td>27,394</td>
<td>29,949</td>
<td>9.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backward Indirect Impacts (Million €)</td>
<td>1,189</td>
<td>1,269</td>
<td>6.66%</td>
</tr>
<tr>
<td>4.3</td>
<td>Incidence of the Data Economy on GDP (Only direct and backward indirect impacts)</td>
<td>Ratio between value of the Data Economy and GDP (%)</td>
<td>0.93%</td>
<td>0.95%</td>
<td>1.59%</td>
</tr>
</tbody>
</table>

Source: European Data Market Monitoring Tool, IDC 2018

International Overview and Comparison with the EU

In line with the results obtained in the previous European Data Market Study (SMART 2013/0063), the U.S. continue to enjoy the highest impact of the Data Economy on their GDP - 0.81% in 2017, up 4.9% with respect to 2016.

While not leading in absolute values, Europe emerges as the most dynamic region with a sustained and unsurpassed impacts’ growth of more than 9% year-on-year 2017 at the level of the EU27. If the U.K. is added to the equation, the incidence of the Data Economy would be more than 0.5% of the EU28 GDP – a double-digit growth with respect to the previous year. Europe thus represents a growing and dynamic Data Economy: in terms of size and growth, the value of its Data Market (as defined by the European Data Market Study) is second only to the U.S.; more interestingly, the impact that this market generates on the economy as a whole (the “Data Economy”) has become more and more visible over the past few years (2014 through 2017) thus rapidly catching up the gap with the American economy.
Interestingly, though, Europe appears to be catching up on its gap with the U.S. and shows a renovated dynamism in some of the most significant Data Economy’s areas. In terms of data suppliers, for example, the EU can exhibit a year-on-year growth 2017-2016 of 9.2% - more than twice than in the U.S. and three times stronger than in Japan over the same period.

While close to double-digit growth, the value of the Data Market in the EU28 has marked a relative halt with respect to the U.S. in the period 2016-2017 but has continued to largely outpace Brazil and, to a much lesser extent, Japan.
Notwithstanding its dynamism, the EU continues to suffer from higher levels of fragmentation: Europe is more divergent than the U.S. when it comes to the usage of digital technologies across companies, including those in the same sector. This is partly due to structural factors (such as the higher presence of SMEs in Europe than in the U.S.) and to cultural and educational factors (such as Europe’s relatively weaker position in creating and keeping the necessary digital skills to support the digital transformation process).

Nevertheless, the EU remains a protagonist in many areas of the Data Economy - it has a lively digital start-ups scene and a considerable innovation capacity, but it is still unable to translate this potential into global digital platforms as it is done in the U.S. - bar few notable exceptions, to be fair, such as Spotify in Sweden, Deezer in France, Shazam in the U.K. and Gemalto in the Netherlands, just to mention a few. In terms of research & innovation, and of the subsequent innovative technologies, Europe is slower the U.S. in building an effective ecosystem to turn these technologies in commercially exploitable applications. In areas such as Robotics, Augmented Reality/Virtual Reality and machine learning for instance – all future engines of the digital transformation – companies like ABB, BMW, Bosch and Siemens are investing heavily but find it difficult to build a critical mass competing at a par with American manufacturers.
3. DESCRIBING THE DATA MARKET – THE QUANTIFIED STORIES

Two quantified stories were produced to accompany the First Report on Facts & Figures (D2.1) and the First Report on Policy Conclusions (D2.2). The stories focused on the use of data produced by the private sector and by scientific research to address public issues. The research investigated the drivers and inhibitors currently at play in Europe to open up private data to better serve the wider public interest; it further examined the dynamics underpinning (or hindering) the sharing of scientific data to support innovation in Europe. A third quantified story on data monetization is currently under development. The aim of this research is to investigate how companies and organizations can generate and measure economic benefits by leveraging data generated through business operations, publicly available data or data collected via electronic devices and sensors (i.e. IoT). Given the slow pace in obtaining interviews to feature real-life examples of data monetization at European level, the research is ongoing and will be completed in July-August 2018. The results of the third story will therefore feature into the next edition of this report (D2.6 Second Interim Report) to be published in March 2019.

3.1 Opening Private Data and Scientific Data for Public Interest

The research conducted by the study team⁷ presented a series of initiatives focusing on the leveraging of private data for public interest – either upon the impulse of companies, such as data philanthropy and data grants, or by the initiative of public Statistical Offices, such as public private partnership, procurement and regulation. The analysis revealed that most initiatives take place in developing countries, where the lack of reliable statistics calls for the identification of alternative proxies. Many different data types and application domain can be found, although the most widely used dataset is certainly mobility data from mobile phone providers. The application domains vary widely, from tracking health and epidemics to measuring tourism or poverty levels. Indeed, there is scope for both disruptive innovation (e.g. reduce the mortality rates of malaria, legionnaire and Ebola) and incremental innovation (e.g. reduce the cost of producing statistics). The research exposed the main advantages of using privately held data, that is:

- Reducing the costs of data collection for statistical offices, and the administrative burden, by using already available data rather than surveying companies and individuals;
- Making new data available, through the provision of real-time and extremely granular data for instance over population mobility;
- Removing bias deriving from self-reporting data in favour of observational data. The limitations of self-reported data are well known and strongly limit our understanding of human life.

At the same time, privately held data are not a low hanging fruit. They cannot be immediately used as a substitute of official statistics, typically because of their biased coverage. They need significant resources to be used, and while there is evidence about the cost savings, the impact in terms of large scale disruptive innovation remains incipient. Indeed, in most cases projects are still at the level of feasibility or pilot. Companies are reluctant to share data not only because of a legitimate concern over giving up sensitive data, but most often because of concerns related to privacy.

infringement risks, related to the costs and risks of anonymisation techniques. Costs and skills are significant barriers on both the supply and the demand side.

To achieve these benefits, and address the related challenges, the story highlighted that companies typically use a set of organisational arrangements, depending on the specific context:

- Data Cooperatives or Pooling;
- Prizes and Challenges;
- Research Partnerships;
- Intelligence Products;
- Application Programming Interfaces;
- Trusted Intermediary.

Overall the research revealed that companies are more likely to exchange data when they are aggregated and shared only with an individual organisation rather than as open data. Depending on the purpose of the exchange and on the type of data in object various sharing agreements and different sharing techniques can be chosen. It is worth notice that when data are shared with public entities they usually contain a higher level of information with respect to when they are opened up to everyone. However, trust in institutions and the consequent re-use of the information matters.

The choice of the most suitable data exchange arrangement, and the decision over the kind of policy measures to be adopted to support the arrangement, depends on several factors:

1. The benefits that can be drawn from data. For instance, one can differentiate between emergency situations related to natural disasters, security and epidemics, which could justify stronger policy measures, against data necessary to monitor tourism.

2. The role of government in the data gathering. In regulated industries, or in cases of procurement or concessions, it is easier to justify a regulatory approach.

3. The costs of data sharing. For instance, real time, disaggregated anonymised data is far more costly than aggregated bulk data. The higher the costs, the more difficult it is to justify companies’ obligation to open their data.

4. The commercial sensitivity of the data. Typically, data about customers is the most commercially sensitive data for companies, and it is more difficult to obtain than other data.

3.2 Story 2 – Opening Scientific Data for Innovation

The second story investigated how an enhanced availability and sharing of scientific data could benefit the development of innovation. While there is little doubt that opening up scientific data for reuse will increase the quality and productivity of science, still little is known about the implications for innovation. The limited literature in this domain is not conclusive and points out to both positive and negative effects: some scholars predict openness will negatively affect business commercialisation of research outputs. Others assume greater industrial collaboration due to greater availability of and access to scientific outputs. What is clear is that there is a tension between the different goals set to research organisations and there are potential trade-offs between these goals.

---

8 D3.2 “Opening Scientific Data for Innovation” Update of the European Data Market, SMART 2016/0063
The story featured the case study “Open Targets”, which was chosen because of its pioneering nature, in a domain heavily driven by IP such as the pharmaceutical industry. It showed that, driven by the need for increased innovation, open data sharing happens even in one of the most heavily protected research-market domains, provided it is carefully designed through a combination of discriminatory and open access. The “Open Targets” project includes a data sharing platform between companies and research centres, as well as the later public release of open data for reuse after a time lag. It has achieved substantial uptake and impact, both in scientific and innovation terms.

Our analysis revealed a tension between the goals of scientific openness and commercial exploitation of research results. Yet this tension does not imply an incompatibility between the goals, but the need to identify smart solutions that adequately balance the different needs. Indeed, open scientific data are a reality and a source of innovation even in the most IP-intensive domains, such as pharmaceutical. In other words, companies are reluctant to share scientific knowledge and data, but can accommodate such openness if justified by the quality of the research partnership, if still developed in a pre-competitive phase in research and if protected by provisions such as delayed publication periods. Based on this, an initial set of policy conclusions could be summarised as follows:

- **There is not such a thing as one-size fits all.** Opening up scientific data varies a lot over different disciplines and might affect differently innovation and future commercialisation of future products and services, depending also on sector characteristics (e.g. there are industries more sensitive to IP for future R&D investment than others).

- **Promote smart openness to foster innovation.** It is important to move away from the ideological debate that confront the pro-Open with the pro-IP advocates, and find the appropriate boundary conditions and frameworks to opening scientific data, which has undoubtedly a positive effect to scientific progress and society, while not threatening or compromising innovation and industrial collaboration from research infrastructures and companies.

- **While peers and journals are the main motivators, the Government can play a seductive and persuasive role.** Governments have a wide range of possibilities when trying to foster open scientific data (e.g. regulation, funding, awareness etc.). However, according to recent studies in this matter (Kim, 2017) due to the small impact of governmental policies on scientists’ behaviour when sharing data, it would be more effective if convincing intermediary players (e.g. journals and stakeholders in the scientific community) to promote and engage in opening up scientific data. Direct policy from the governments would be not so effective like advocacy policies to other stakeholders in scientific community for opening up scientific data. In the long-term, governmental policies should try to impact the incentives system of scientists in order to make them more interested in sharing data besides the open inquiry spirit of science.

- **Data publication is the beginning, not the end of public-private collaboration.** Uploading scientific data is not enough in order to foster its re-use and to translate it in innovative products and services. Close work from scientists who generate the data and the researchers (in companies) that seek to re-use it is still needed to understand what data tells, how it was generated and how to interpret it. The Open Targets case highlights the importance of working with companies and users of the data to understand how to integrate, make the search, visualize data, workloads etc.

- **Increased monitoring.** There is a lack of evidence about the impact of scientific data re-use to innovation. This is due to several methodological challenges. However, we need to overcome such challenges, and provide empirical studies that measure the impact of opening up scientific data to innovation performance of companies.
3.3 What the Quantified Stories tell us so far

A common finding across those sectors is that in most cases data per se do not have substantial economic value. The metaphor of data as the new oil is misleading. To generate value from data, substantial effort and high-level skills are necessary, typically requiring the collaboration of those who gathered the data in the first place. In other words, data are rarely exchanged as such, through simple monetary transactions, but typically require either the collaborative effort of the data holder and the data re-user or an open data sharing. This is one of the reasons why data marketplaces have struggled to emerge.

It is therefore clear that the free flow of data can generate maximum societal and economic benefits. There are indeed plenty of projects and initiatives to promote data sharing both in the private sector and in research, although they are far from becoming “mainstream”. What is still unclear is how this free flow of data can be achieved at systemic level. Most of all, there is a clear need to understand the structure of objectives and incentives for different stakeholders. Pursuing the public good is not a sufficient driver for data sharing. Companies are profit driven. They share data typically by selling integrated analytics services, for instance Vodafone offers packaged services to government based on the mobility data gathered by their antennas. In special circumstances, typically in developing countries, they offer their data services as an alternative to poor-quality official statistics, and in this case the main incentive lies in corporate image – and the potential indirect business benefits. Companies also can provide different levels of access under freemium business models: for instance, aggregated data for free and granular data for a premium, as in the case of BBVA Data & Analytics – a data supplier company established by the Spanish bank BBVA, which offers data services to specific private industries (such as retail and tourism) against a fee but also adopts open data access models where interest third parties can get hold of a (limited) amount of data for free through a set of open APIs.

Similarly, researchers are driven by the recognition of their peers. Unfortunately, at this stage data sharing does not provide direct benefits in terms of reputation – for instance, in terms of impact factor or citation rates, which could improve the career opportunities. On the other hand, companies involved in collaborative research projects are reluctant to allow data sharing, in order to maintain competitive advantage. For this reason, a wealth of ad hoc solutions have been designed to preserve some form of competitive advantage for both the researcher and the companies – typically in the form of delayed data publication.

In conclusion, data held by private companies and scientists can be invaluable for addressing societal issues, or for generating new products and services. But they are not a low hanging fruit: they require substantial investment, in some cases with the direct involvement of those who gathered the data in the first place (such as mobile operators). And any mandatory data sharing measure needs to be carefully designed in order not to hurt the emerging Data Economy: each ecosystem is building its own set of business models and organisational arrangements to fit their peculiar system of incentives: from PR to new revenues stream for companies, from increased scientific productivity to career opportunities for scientists. What is clear is that data sharing and reuse are necessary, but there is no “silver bullet”, no easy single solution to ensure it. It is a matter of designing the right policy mix of raising awareness about the different modalities, building skills, and removing existing barriers.
4. MAPPING THE DATA MARKET – DATA LANDSCAPE AND DATA MARKET MONITORING TOOL

4.1 The EU Data Landscape

The first new EU Data Landscape review (Deliverable D4.1) was performed in January 2018 and is expected to be further updated in January 2019 (Deliverable D4.2). At the time of the review, data landscape database included a total of 1256 companies and covered 36 countries (EU-28, Belarus, Bosnia and Herzegovina, Iceland, Moldova, Norway, Serbia, Switzerland and Ukraine). By the end of May 2018, the situation did not substantially change, with 56 new companies added to the database between February and May 2018.

Among 1256 companies, 94 have been identified as key data landscape companies in line with a set of criteria adopted.

Major changes to the EU Data Landscape

The new EU Data Landscape review introduced some changes to the approach:

- The methodology has been revised and new categorisation was introduced in order to capture new trends and align the EU Data Landscape categorisation with the EU Data Market Study one;
- The existing database was validated and further extended in geographical and coverage scope. This included the input on big data SMEs from Big Data Value eCosystem Project (BDVe) project;
- The list of key data landscape companies was reviewed according to updated criteria;
- The additional filtering options in an online database: SMEs and key data landscape companies, were introduced.

![Database of Data Landscape companies (wwwdatalandscape.eu)](http://datalandscape.eu/companies)

Figure 11: Database of Data Landscape companies (www.datalandscape.eu)

Overview of the EU Data Landscape Database (status in January 2018):

- **783 additional companies** were added to the database (an increase of 65.5% as compared to 473 companies in the previous update of the database).
• **UK companies account for 23% of the total database**, followed by Spain (13%), Germany (9%), France (7%), Netherlands (5%) Belgium (4%).

• **Analytics becomes the biggest category (42%)** growing by 15 percentage points from 2016 - a result of the growing number of AI companies in the AI and machine learning sub-category (394 analytics companies were identified as AI focused).

• The share of companies categorised as **Vertical Applications** diminished (from 27% to 13%) though a number of companies in Vertical Applications grew by 32% in absolute terms. 19% of vertical applications are in the fields of Professional services followed by Information and communication (17%) and financial services (12%),

• **Enabling players category** gained 14 percentage points (due to the introduction of numerous new accelerators, incubators, venture capital funds, business angels and startup associations) becoming, as a result, **the second biggest category** (representing 26% in 2018 against 12% in 2016).

Figure 12: Percentage of Key Data Landscape companies in 2016 and 2018 – top 10 countries

Source: European Data Market Study, D4.1 EU Data Landscape, Review at January 2018

**Key Data Landscape companies (status in January 2018):**

• **Key Data Landscape companies** were selected from the main database according following criteria: 1) The company is listed in the Global Big Data Landscape map,9 or 2) a company received over 1m Euro in funding according to Crunchbase database, and 3) the company has its main headquarter or R&D department in Europe.

• The list of key data landscape companies grew from 68 to 94.

• **UK is leading the list (30%)** followed by Germany (14%) and France (14%).

- **Vertical Applications** remains the predominant category (40% in 2018), followed by **Analytics** (39%) and **ICT Enablers** (17%).

4.2 The European Data Market Monitoring Tool

The new improved version of [European Data Market Monitoring Tool](http://datalandscape.eu), together with the presentation of the new forecasts, was launched on the datalandscape.eu around 20 April 2018. The scope of data was extended from 2013 -2020 to 2013 – 2025 forecasts.

The improved version of European Data Market Monitoring Tool presents the new approach to visualisation:

- A new visualisation of European map was introduced.
- A new divergent, gradient colour scale was adopted in order to highlight the difference between minimum and maximum values (dark orange standing for minimum and dark green standing for maximum). The same scale was adopted for almost all indicators.
- The data tables (on the right-hand side) were elaborated with visualisations of indicators as well as an option to extend the data, that allows a deeper dive into data, and more precise and more advanced analysis.

![Figure 13: The European Data Market Monitoring Tool](http://datalandscape.eu/european-data-market-monitoring-tool-2018)
• A new bubble chart was introduced below the map chart allowing for personalisation and modification of graph.

Figure 14: The European Data Market Monitoring Tool – bubble chart


• Finally, the line charts with the same scales were introduced for Industry indicators so that they are comparable.

Figure 15: The European Data Market Monitoring Tool – Industries

5. ACTING UPON THE DATA MARKET – THE ROLE OF POLICY

The sizing and forecasting exercise carried out by the European Data Market Monitoring Tool, together with the additional analysis obtained through the quantified stories, were complemented by the First Report on Policy Conclusions (D2.2), which investigated the role of policies in shaping the present and future trends of the European Data Market and Data Economy.

5.1 The Role of Policy and the Future of Europe’s Data Economy: The Three Scenarios

The 2025 scenarios developed for Update of the European Data Market Study (SMART 2016/0063) present potential evolution paths of the European Data Market and Economy, driven by different macroeconomic and framework conditions, highlighting the critical turning points to be faced in the next years by governments, businesses and social actors. The scenarios presented here are focused on the year 2025, taking as a reference starting point the 2020 scenarios presented in February 201710. They are built around the intersection of two main focal issues:

- **The Data Market’s pace of growth**: how fast will data-driven innovation grow in Europe? The scenarios outline the 3 alternative possibilities of slow, medium or fast pace of innovation.
- **The potential evolution of the model of data governance**, in terms of how the ownership, access, control and exploitation of data assets will be managed. To put it more bluntly: who will have power on data and what will governments do about it? The scenarios outline future models ranging between two potential extremes: on the one hand, a data governance model where a few data holders (private or public) control most of data assets; on the other hand, an open and participatory data governance model, based on sharing and transparency.

The Data Economy scenarios therefore are positioned at the intersection of these two main focal issues as follows:

- **The Baseline scenario** is characterised by a healthy growth of data innovation, a moderate concentration of power by dominant data owners with a data governance model protecting personal data rights, and an uneven but rather wide distribution of data innovation benefits in the society. This is considered the most likely scenario.

- **The High Growth scenario** is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society;

- **The Challenge scenario** is characterised by a low level of data innovation, a moderate level of data power concentration due to digital markets fragmentation, and an uneven distribution of data innovation benefits in the society.

The scenarios explore the drivers and framework conditions which may lead to maximise the benefits of a balanced Data Economy and to avoid the risks of an unbalanced one, highlighting the consequences of policy actions.

---

Policy and the Baseline Scenario

The Baseline scenario predicts that the current positive economic climate and growth dynamics of the European Data Market will continue towards 2025, driven by a healthy growth of the European data industry, a continuing improvement of the offering of data products and services, and a corresponding gradual development of demand, especially by the most advanced, competitive and innovative enterprises, large and small. Nevertheless, leading companies and regions will increase their advantage as more traditional companies and sectors struggle to move at the same speed. Another potential factor could be the uneven diffusion across Europe of 5G infrastructures needed to support next generation online services and especially IoT.

In the Baseline scenario, policy plays a mitigated role with mixed success. Today the European Data Market is fragmented, and the diffusion of data-driven innovation is uneven, but we are on the verge of introduction of several policy measures aiming at building the Digital Single Market, first of all the General Data Protection Regulation (GDPR) and the Free Flow of non-personal Data initiative (FFoD). According to IDC surveys, over half of EU enterprises are struggling with the implementation of GDPR aiming for pragmatic compliance, while very few are looking at it as an opportunity. We foresee the GDPR to create gradually a successful harmonisation of regulation across the EU, but we suspect that organisations will need a long period of adaptation. Removing barriers to the flow of non-personal data across Europe is a critical success factor to unlock the exploitation of European datasets at a scale and scope sufficient for the new data-driven processes such as machine learning. This is the aim of the FFoD initiative, whose proposal for a Regulation however is still in the legislative phase between the European Commission, the European Parliament and the Council. In the best possible perspective, the FFoD Regulation will be adopted by 2019 but will take a few more years to take effect, spreading across Europe at variable speed. In short, the completion of the Digital Single Market will progress and possibly be completed only at the end of the forecast period. On the other hand, in this scenario R&D investments in Big Data through Horizon 2020 in collaboration with the contractual Public Private Partnership on Big Data will sustain the deployment of data-driven innovation, and the deployment of Digital Innovation Hubs such as the Big Data Centres of Excellence will stimulate industry-research interaction and promote the supply of data skills.

Policy and the High Growth Scenario

In the High Growth scenario, the growth of the European Data Market will enter a faster growth trajectory than in the Baseline scenario and the adoption of data technologies will spread beyond the minority of pioneers to a wider population of mainstream users. Digital transformation, data monetization, B2B data sharing on multi-user cloud platforms will spread faster than in the Baseline scenario. All industries will keep pace, also the public sector, even though the intensity of data innovation will grow faster in Finance, Manufacturing, Professional services, ICT and Media. The supply-demand dynamics will change from technology-push to demand pull, with a fully developed ecosystem generating positive feed-back loops between data companies and users.

European policies have a relevant role to play in this scenario. The completion of the Digital Single Market within the forecast period is a key success factor: this requires rapid and successful implementation of the GDPR and the Free Flow of non-personal Data initiative, with true liberalisation of data flows across Europe. R&D investment at EU and national level must be particularly effective and the Network of Excellence Centres must succeed in becoming innovation beacons in every region as well as increasing the supply of scarce data skills. In short, in this scenario Europe captures in full the digital opportunity, as advised by McKinsey.
Policy and the Challenge Scenario

In the Challenge scenario, European innovation forces become lost in a maze of barriers and are unable to overcome them, resulting in much slower Data Market and Data Economy growth. This is a more extreme version of the Baseline scenario, where markets fragmentation and the failure to complete the Digital Single Market exacerbate regional and industry differences, so that only the best enterprises and the richest regions keep pace with the technology race. This scenario foresees a negative self-reinforcing circle, where less positive global economic conditions discourage investments and weaken global demand with a negative impact on European growth. A slower pace of digital innovation deprives the economy of the boost to growth potentially given by data-driven services and products, while enterprises find competing in international markets more difficult.

In policy terms, this scenario is driven as much by the failure of the Digital Single Market and of innovation investments than by global economic trends. A missed approval or weak implementation of the Free Flow of non-personal Data would result in the persistence of digital barriers. The lack of increasing investments in AI, robotics and automation would create a disadvantage for the European industry.

5.2 The Role of Policy in the Era of Digital Transformation

The European Data Market, its evolution and impacts on the overall economy and society are closely related to the process of digital transformation (i.e. the re-organisation of business processes to capture the value of digital technologies), which, at its core, has companies and organisations across the whole of the data value chain – from those creating and collecting data, to those storing, aggregating, curating, analysing, marketing and distributing them, to those using or reusing the data. For this reason, since its inception in 2014, the European Data Market Study has carefully defined and apprehended the “data companies” phenomenon by measuring both data suppliers (that is companies producing and delivering data-related products, services, and technologies) as well as data users (companies exploiting and analysing data intensively and using improve their business). According to the Data Market Monitoring Tool the number of data companies has been incessantly increasing over the past four years; what is more, IDC’s independent research has shown that a clear majority of European organisations are evolving in their digital transformation journey and are starting to see a pay-off from their digital efforts and generating positive returns. Figure 13 below shows that the relative majority of European companies (29%) is already a “digital player” according to IDC’s definition of the digital journey – a percentage that becomes an absolute majority if the category of “digital explorers” is added.

*Figure 16: The Digital Transformation Journey in Europe (2017)*
The trend outlined above is expected to consolidate and continue unabated in the few coming years with the addition of cognitive and AI (Artificial Intelligence) technologies to the equation of digital transformation. Indeed, the exponential growth of data requires intelligence to be part of every business model, product, solution and services.

European companies find themselves in a critical stage amid this unrelenting growth of the digital transformation process with several policy challenges lying ahead. The war on talent, for example, is heating up. Europe has long faced a skills gap for ICT skills, but the digital transformation process is also adding the need of new digital skills to operational and business staff. Increasingly, these employees will be expected to have greater analytics and AI competences in order to be able to use and make sense of large data sets to improve and automate process tasks. As it will take time for the European education systems to adapt to the growing demand for these combination of skills, policy intervention will be required and organisations will need to consider a combination of internal training programs for existing staff and close cooperation with educational institutions to ensure graduates will have the required skill sets. Similarly, Europe’s evolving regulatory environment around data privacy, protection, and security will have to be closely watched to reduce possible effects and delay on the impact and timing of data monetization compared with other parts of the world. Indeed, some European industries are highly regulated – this is the case of Manufacturing and Healthcare, for instance. This, combined with a fragmented, country-based IT landscape, could inhibit the level of data access, security and governance necessary to drive a successful cognitive/AI project. Again, appropriate policy adoption and enforcement is needed to go beyond Member-State specific use cases and leverage the value of existing data across multi-company ecosystems.

5.3 Policy in the Data-Driven Society

Data-driven technology has allowed the transformation of data into assets and the emergence of data-driven services providing benefits to consumers and citizens. In fact, the positive impacts of data-driven innovation go beyond productivity growth, as it can also directly contribute to inclusiveness, development and the well-being of citizens. In particular, science and education, healthcare services and public administration have been identified by the OECD as the sectors where data-driven innovation could lead to the highest impact.

However, there is also increasing concern about the new ethical and social issues arising from the diffusion of Big Data which go much beyond the debate on data and privacy protection. On the one hand, there is concern about the concentration of power in the hands of a very small number of leading data platforms. For instance, Facebook with its 2.2 billion users dominates the global “social graph”, which means most of our social interactions. Google dominates as much as 85% of online-search-ad revenues worldwide. Amazon is a retailer as well as a marketplace and controls some 40% of America’s online commerce. In the last months, Google, Facebook and Amazon have come under fire for being what the Economist defines as “BAADD”: too big, anti-competitive, addictive and destructive to democracy. These are only the most visible trends. Another relevant issue, not easy to grasp by the general public, is algorithmic transparency11. The combination of increased availability of large amounts of fine-grained human behavioural data and advances in machine learning is presiding over a growing reliance on algorithms to address complex societal problems. Algorithmic decision-making processes are being used in the U.S. to give credit, define the level of insurance or

---

banking fees, or even to keep people in jail. These systems may lead to more objective and thus potentially fairer decisions than those made by humans who may be influenced by greed, prejudice, fatigue, or hunger. But data scientist Catherine O’Neill has shown how this use of decision-making algorithms without transparency and redress mechanisms may have discriminatory impacts and generate information and power asymmetry. With the progress of Artificial Intelligence, even the owners of these algorithms often do not know how and why decisions are made, never mind being able to identify mistakes and fix them.

5.4 Europe’s International Competitiveness and the Role of Policy

The international indicators measured by our latest Data Market Monitoring Tool draw a mixed picture: the EU still lags behind the U.S. in terms of size of the Data Market and the Data Economy but has by no means reached its potential from digital technologies. Europe’s relative delay in keeping up with the U.S. in these strategic areas may have undesired consequences on the employment and the future of work in the EU. While technologies such as robotics and machine learning may pose a threat to the current level of employments, other digital technologies may help create new opportunities and more flexible job allowing displaced workers to recoup income as independent workers. In Europe, though, only about 15% of independent work is conducted on digital platforms versus 30% in the U.S. - the proportion is increasing and bears good hopes for the future but also rises questions such as dissatisfaction with income variability, lack of benefits associated with traditional work. These issues will have to be tackled sooner than later in Europe to avoid further social disruption.

Indeed, Europe’s competitive position against the U.S. will not improve until all the measures in the Digital Single Market Strategy are implemented across all EU countries. Self-driving cars, for example, are starting to roam American streets under test conditions and the U.S. government has clear policy recommendations on this area. The EU has a less coordinated approach, which brings to the fore the importance of the DSM and the need to remove the barriers that still impede the free flow of data within the EU both at cross-border and at intra-company level. The end of geo-blocking by the end of 2018 is certainly a very good step in this direction. European companies, though, should also be adequately supported and incentivised to open up and share their data. Indeed, data users and data suppliers should be put in a position to scale within and beyond individual Member States and realise the potential of an open and effective Digital Single Market.

On the other hand, a significant increase in digital infrastructure investments and digital skills should be carefully considered to deepen and expand Europe’s digital ecosystem. In this respect, particular attention should be devoted to Europe’s already vibrant next generation start-ups ecosystem. Many good practices and excellent examples abound here (Berlin, Amsterdam, Barcelona, London, Stockholm, Paris to name a few top-level European cities), but more should be done to experiment and embrace innovative digital technologies such as AI, machine-to-machine, robotics and others – the real drivers of the digital economies in the upcoming years.

Finally, the current efforts undertaken at national and EU level towards a significant improvement in education and skills should be pursued and step up. Worker displacement and transition would be inevitable as digitisation progresses and policy makers and business leaders alike should pay more attention and investments into Science, Technology, Engineering and Math (STEM) skills through

---

school systems, while favouring effective retraining for mid-career workers as these transition to the future-of-work-like positions.
6. CONCLUSIONS

While digital transformation continues unabated, Europe has entered 2018 with a new momentum backed by moderated economic growth and a general improvement of most of the indicators that have been measuring the state of the European Data Market and the European Data Economy over the past three years.

Indeed, the overall development of the six interrelated indicators apprehending Europe’s Data Economy in 2016-2017, and through the year 2025, can be seen holistically along four main dimensions:

- The Workforce and Skills dimension including the measurement of data professionals and their potential skill gap.
- The Supply and Demand dimension incorporating the measurement of data supplier and data user companies and the revenues generated by these companies.
- The Business and Economy dimension comprehending the size of the Data Market and the value of the Data Economy.
- The International context dimension including a select number of indicators for Brazil, Japan and the US.

*Figure 17: The four Dimensions of the Data Market’s Key Facts & Figures*

Source: The European Data Market Monitoring Tool, IDC, 2018
6.1 Data Professionals

Data professionals are estimated at 5.3 million in the EU27 and 6.7 million in the EU28 in 2017, thus marking a significant increase in 2017 over the previous year (8.2% and 8% year-on-year respectively). Even more interestingly, the employment share and the intensity share components of the data professionals’ indicator are also expected to significantly improve in 2017 and 2020 if compared to our estimates in 2016 (now estimated at 3% and 3.4% in 20017 and 2020 in the EU27 and 3.2% and 3.6% for the same years in the EU28), thus confirming the positive evolution of the workforce involved in data-related professions over the period under consideration. This result shows direct continuity with the findings obtained in the previous European Data Market, SMART 2013/0063. Furthermore, the number of data professionals in the EU28 is expected to rise by a compound rate of 6.7% by 2025. This growth could be as high as 10.1% per year if the highest growth scenario is achieved – giving nearly 13.5 million data professionals in the EU28 and almost 11 million data professionals in the EU27 at 2025.

In terms of data professional skills, a structural imbalance persists between demand and supply of data skills in Europe since the first measurement of the EDM Monitor for the year 2014. In the year 2017, given the strong increase of demand of data professionals from 2016 (+7%), the estimated gap grew by 10% reaching approximately 449,000 unfilled positions, corresponding to 6% of total demand. By 2020 we expect the gap to expand to 699,000 unfilled positions in the EU27, corresponding to 8% of total demand. The 3 scenarios highlight the diverse potential trajectories of the demand-supply balance of data skills to 2025, predicting an increasing data skills gap, ranging from 10% of demand in the Challenge one, to 11% in the Baseline, to 19% in the High Growth (EU27). The slightly lower dynamic of data skills demand in the U.K., because of Brexit negative consequences, means that the gap incidence is smaller in the EU28.

The absolute size of the data skills gap is also relevant, potentially reaching 1 million unfilled positions in 2025 in the EU27 Baseline scenario but up to over 2 million in the EU27 High Growth scenario. This raises difficulties for policies and initiatives aiming at addressing the gap. The gap forecasts extrapolate current supply and demand trends. Actions to improve supply and make the market more efficient, reducing mismatches, can potentially reduce these gaps. The insufficient provision of data skills is a challenge particularly for the data industry which needs to hire highly specialised, technical data skills more difficult to find.

6.2 Data Companies

In the area of data companies, the number of data suppliers continues to grow at a faster pace than the number of data users: the former is estimated at almost 142,000 in the EU27 and at more than 276,000 units in the EU28, thus exhibiting a year-on-year growth of 5.7% in 2017. Data users, instead, are projected to grow at 2.1% in 2017, amounting to almost 516,000 in the EU27 and to 690,650 units in the EU28. If compared to the measurements carried out by the European Data Market Monitoring Tool over the period 2013-2015, these latest estimates show a more dynamic pictures of data companies in the EU, with growth rates constantly increasing over the past four years.

Data users’ penetration rates (i.e. number of data user companies on total companies) vary significantly across Member States and present a wider range than the penetration rate exhibited by data professionals (i.e.: number of data professionals on total employment). With the EU28 average standing at 6.6%, data user’ penetration is as low as 1.0% for Romania, and as high as 12.2% for the Netherlands. This relates to the mix of industries associated with each country: those countries with
a greater preponderance for data-oriented industries show a higher penetration and will continue to do so. Adoption rates of data technologies are higher in industries such as Professional services, Retail, and Financial services.

Data companies’ revenues (corresponding to the aggregated value of all the data-related products and services generated by Europe-based companies, including exports outside the EU) account for 3.2% of total company revenues in 2017, as companies continue their transition to the digital economy. Data revenues are expected to follow the Data Market, as imports and exports of data tools and services tend to follow each other. Forecasting data companies’ revenues shows an expected annual growth rate out to 2025 of 8.3% - easily outpacing the growth of the total ICT market over the same period (expected to be 1.6% from 2016 to 2025 Baseline). The smaller Member States show the highest long-term growth as they have a smaller base from which to grow, but the larger Member States will make the biggest overall contribution to the Data Economy out to 2025.

6.3 Data Market

The European Data Market - i.e. the marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data – is estimated to grow by 9.3% in 2017, and at an average rate of 6.0% out to 2020. This makes the market valued at close to 77.5 Euro Billion with the vast majority of Member States showing strong growth, well ahead of the expected growth for the IT market as a whole, which is projected to grow only 2.6% in 2017, and at an annual rate of 1.7% to 2020.

The larger industries account for the largest share of the Data Market – as there is the greatest number of companies within these industries, but adoption rates of Data Technology tend to be high in Finance, Professional services, and in Retail, which, together with the size of these industries makes them the biggest consumer of data technologies. Manufacturing’s sheer size in the EU economy makes it the largest industry in the Data Market. However, there is significant scope for increased adoption of data technology in the manufacturing industry, so its leading position is unlikely to change. The forecast for the Data Market shows that industries like Manufacturing, Finance, Professional services, Information Technology, and Retail account for more than 75% of the total market growth from 2016 to 2025, with more than half coming from the three main industries of Manufacturing, Finance, and Professional services. The Data Market will continue to out-grow the total ICT market, with its share of this market rising from close to 10% in 2016, to more than 15% by 2025.

6.4 Data Economy

The Data Economy – measuring direct, indirect, and induced effects of the Data Market on the economy as whole – continues to show solid growth across all Member States. Direct impacts will continue to grow and mirror the positive developments of the overall Data Market; the indirect impacts will be the key impacts in all the scenarios, while induced impacts at the same time will start increasing, as the effects in the economy, through job and revenue creation as additional impacts will start to be visible.

In the Baseline scenario the indirect and induced impacts will be more balanced than in 2017 and 2020. The overall impacts for EU27 will be 4.0% of the GDP. The indirect impacts will still play a fundamental role, as the user industries will consolidate the quantitative benefits from the use of data, but more than that, these benefits will go beyond the users and will translate in higher induced effects, generating jobs and revenues beyond the data companies itself. With respect to the
previous measurements, from 2020 and 2025 there will be a significant change in the share of indirect impacts across the three scenarios. In particular, in the High Growth scenario the penetration rate of data products and services into the user industry will be lower, and replaced by higher induced effects. The positive conditions under the High Growth scenario will lead the overall impacts to exceed 1 trillion Euros in 2025.

In terms of industry, the Financial sector, Manufacturing, and Professional services industries will all be heavily impacted by the exchange of data-driven products and services. Indeed, as also shown in the Data Market estimate, these industries make a significant usage of data-related technologies, their forward and backward impacts are high, and can convey effects at an induced level more than other industries thanks to several trends, such as the diffusion of IoT and Cloud computing, the digital transformation processes, as well as the usage of mobile and social technologies.

6.5 The Data Economy beyond Europe – Brazil, Japan and US.

This report extends the analysis of the Data Market and the Data Economy to three main extra-European countries: The United States, Brazil and Japan.

The U.S. confirm to have the healthiest Data Market and Data Economy with consistent and sustained growth in the number of data professionals, companies, and the overall Data Market. Brazil shows lower growth following three difficult years for the country, where investment in the Data Market took a lower priority. However, it is showing signs of recovery and has the potential for significant growth over the next few years. Japan’s market is the closest to the European one in terms of growth and investment. While regaining momentum, Japan’s economy is still hampered by weak internal demand and lack of consumption. This, in turn, produces negative effects on ICT spending and, as a result, hampers the Data Economy and the Data Market potential.

According to our first round of measurement of the international indicators, the European Data Market and Economy in the period 2016-2017 continues to consistently hold second place after the U.S. in value and, to a much lesser extent, in growth. Europe, however, presents a growing and dynamic data ecosystem on both fronts – the Data Market and the Data Economy: in terms of size and growth, the value of its Data Market (as defined by the European Data Market Study) is second only to the U.S.; more interestingly, the impact that this market generates on the economy as a whole (the “Data Economy”) has become more and more visible over the past few years (2014 through 2017) thus rapidly catching up the gap with the American economy.

6.6 The Road Ahead – Plenty of Opportunities and a few Challenges

Europe appears to be set on the right path to reap the benefits of digitisation; however, the actual magnitude of these benefits will depend on EU policies and on the implementation of the Digital Single Market (DSM) Strategy and of its associated policy initiatives. Indeed, the mid-term review of the DSM Strategy in 2017 has shown progress, but also highlighted several policy measures urgently needed to fully unlock the Data Economy, first among them the Free Flow of non-personal data initiative, whose proposal for an EU Regulation is currently under negotiation in the European Parliament and the Council.

Alongside the initiatives directly related to the evolution of the Data Market and the Data Economy, further issues such as the diffusion of open data practices in Europe and the perspective of using private data and scientific data in an open data mode for public interest and innovation should also be paid renewed policy attention to make Europe’s Data Economy progress even further. In this context, other Data Economy-related topics such as the unrelenting growth of digital use cases, the
growing adoption of the “data-as-a-service” model (and the associated phenomenon of data monetization) and the effect of the increasing expansion of data users (especially in traditionally non-IT-dominated sectors and industries) should be further investigated.

The uncertainty currently surrounding Brexit may pose a threat to a balanced development of Europe’s Data Economy over the next few years. Given the strengths of the British Data Market and Data Economy, keeping digital data flowing between the EU and the U.K would crucial both the Britain and for the rest of Europe. Yet, data flows are not officially on the negotiations agenda and British and EU officials have yet to start discussions about how such data flows. Policy makers from both sides of the Channel should therefore devote extra care to this matter with the aim to strike a comprehensive and satisfactory deal allowing both British businesses and EU businesses to access and exchange data under new, mutually acceptable provisions.

Vis-à-vis some of the most significant EU partners on the global scene, policies should be taken into consideration to protect and maintain Europe’s competitiveness. In fact, while the EU still lags behind the U.S. in terms of size of the Data Market and the Data Economy, Europe displays a remarkable potential. However, until all the measures in the Digital Single Market Strategy are implemented across all EU countries its position vis-à-vis the U.S will not significantly improve.

---

13 On this issue please cfr.: “Brexit threat to data flows” by Mark Scott and Annabelle Dickson, 03.09.2018, POLITICO
METHODOLOGICAL ANNEX

Overview

In line with the methodology adopted in the previous European Data Market Study (SMART 2013/0063), the measurement methodology for this updated report was based on the steps outlined in Figure 15 below. Compared to the previous steps it does not include the ad-hoc surveys which were used to establish the baseline. However, thanks to the use of IDC primary research data tracking the market, we have already proven the feasibility of updating the indicators without repeating the initial surveys.

The main steps of the methodology did include:

- Desk research on the main EU and global national and statistical sources; each indicator has specific set of sources;
- Extraction of data from the relevant IDC surveys and databases;
- Additional secondary research and case studies interviews for the stories, which in turn did feed back to the indicator models to help in the modelling and estimate of indicators;
- A selected number of opinion leader and stakeholder interviews to feed into the modelling and scenario assumptions;
- Implementation of the 7 indicators models and elaboration of results;
- Development of the forecast scenario assumptions and update of the 3 scenarios;
- Assessment of policy insights building on the results of the previous steps.

*Figure 18: A sophisticated Methodology*

**Desk Research**

As done in the first study, the study team reviewed the list of relevant public sources and updated it to collect additional relevant data. The list of the main sources used so far is outlined below and will be constantly revised when preparing the second and the third round of measurement of the indicators.
Concerning the indicators on Data Market, data companies, data companies’ revenues, and the Data Economy the main sources were:

- Eurostat business demography statistics in the European Union, treating aspects such as the total number of active enterprises in the business economy, their birth rates, death rates, and the survival rate (last update: December 2014);
- Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: March 2016);
- IDC’s detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 and 2015;
- IDC Worldwide Black Book (Standard Edition), quarterly updates from the years 2014 through 2015. The Black Book represents IDC’s quarterly analysis of the status and projected growth of the worldwide ICT industry in 54 countries.
- IDC End-User IT Trends and Digital Transformation: IDC European Vertical Markets Survey 2015
- IMF World Economic Outlook (WEO) Database, April 2016

For the data professionals we used in addition the following sources:

- European Data Science Academy (EDSA) project deliverables and publications (July 2015).

For the indicator on the Citizens’ Reliance on the Data Market we used in addition the following sources:

- The Digital Economy and Society Index (DESI), Human Capital Dimension, (2a Basic Skills and Usage; 2b Advanced skills and Development), last update, February 2016.
- IDC European Quarterly Wearables Tracker Results: Western Europe 3Q15 Analysis, January 2016

**Ad Hoc Workshop and Expert Interviews**

For the update of the methodology and of the assumptions for the indicators models and the forecast scenarios, the study team has carried out a few selected, one-to-one interviews with key experts and has organised a specific workshop in collaboration with the BDVe project.

The workshop on “The European Data Economy by 2025” was led by IDC on October 20th, at the BDVA’s premises in Brussels. It gathered insights from the high-level group of industry and research experts, from the BDVA community, about the potential growth paths of the European Data Economy by 2025. The group identified and discussed its possible key turning points, and the most impactful drivers and barriers of the data-driven innovation.

Four thematic areas were discussed in the plenary session and in ad-hoc breakout sessions:

- Supply and demand dynamics;
- Technology trends;
- Policy and regulation;
- Social role of data.

A list of 21 different evolution paths of the data economy was identified and a narrative developed for each of these possible paths. A subsequent discussion in the afternoon successfully trimmed down the list to a manageable number of seven fully developed scenarios on the basis of which, the assumptions underpinning the different set of indicators were updated.

In the course of the workshop, the study team also conducted a few selected interviews with key experts to:

For the update of the methodology and of the assumptions for the indicator models and the forecast scenarios, we carried out 4-5 key expert interviews in the first phase of the study focused on:

- Validation of the methodology approach
- Feedback on main factors affecting the dynamics of data-driven innovation
- Suggestions of improvement

Interviews were carried out with:

- Paul Czech, Know-Center GmbH, Research Center for Data-Driven Business & Big Data Analytics
- Anthoine Dusselier, Dawex
- Bas Kotterink, TNO
- Philip Carnelley, IDC

The initial list of experts identified in the proposal phase of the project, and already leveraged during the preparation of the previous study, remains valid and will be used in the upcoming rounds of measurements of the indicators. The list includes:

- Vincenzo Spiezia, Head of the Information and Communication Technologies Unit in the Directorate for Science, Technology and Industry of the OECD
- Jonathan Cave, Senior teaching fellow in economics, University of Warwick
- Elena Simperl, professor of Computer Science at the University of Southampton and manager of the European Data Science Academy (EDSA) project

**Additional Research on Start-ups and DIH**

In the framework of this updated study, IDC has conducted extensive and in-depth desk research on the diffusion of data-driven start-ups and different types of already existing supporting initiatives and entities by relying on a valuable number of primary and secondary resources including, but not limited to:

- Business incubators, accelerators, DIHs own websites and directories;
- European and National Statistic’s Offices’ sources such as Eurostat’s Business demography statistics and other statistics from national offices presenting data on business demography and treating aspects such as the total number of active enterprises in the business economy, their birth rates, death rates, and the survival rate;
• Additional sources from the European Commission such as the Entrepreneurship 2020 Action Plan and the Start Up Europe initiative under the Digital Single Market strategy (https://ec.europa.eu/digital-single-market/en/policies/startup-europe);
• Other sources from the specialized press or the business community such as EU-Startups.com (one of the leading start-up blogs in Europe);
• Ad-hoc studies such as the First and the Second European Startup Monitor carried out by the German Startups Association (GSA) and the European Startup Network (ESN).

For the sake of clarity and completeness, we provide the following definitions of the main types of supporting initiatives and actors identified during the research activity:

**Key Terms - Definitions**

**Business accelerator** A program or organisation characterised by the following common traits: i) an application process that is open to all, yet highly competitive; ii) possible provision of pre-seed investment (grant or equity); iii) a focus on small teams instead of individual founders; iv) time-limited support (usually from 3 to 6 months) comprising programmed events and intensive mentoring; v) cohorts or ‘classes’ of start-ups rather than individual companies.

**Business angel** A private individual, often of high net worth, and usually with business experience, who directly invests part of his or her personal assets in new and growing private businesses. Business angels can invest individually or as part of a syndicate where one angel typically takes the lead role.

Besides capital, angel investors provide business management experience, skills, and contacts for the entrepreneur. Experienced angels also know that they may have to wait for a return on their investment. They can therefore be a good source of ‘smart and patient’ capital.

In many countries, they constitute the second largest source of external funding in newly established ventures, after family and friends.

**Business incubator** An organisation designed to advance the growth of start-ups entering the beginning stages of building their company through an array of business support resources and services that could include physical space, coaching, common services, and networking connections. Incubators operate on an open-ended timeline (usually from 1 to 5 years): they focus more on the longevity of a start-up and are less concerned with how quickly the company grows. Incubators do not traditionally provide capital to start-ups and are often funded by universities or economic development organizations. They also don’t usually take an equity stake in the companies they support.

**Digital Innovation Hubs (DIHs)** Ecosystems that consist of SMEs, large industries, start-ups, researchers, accelerators, and investors, which aim to create the best conditions for long-term business success. DIHs help companies to become more competitive in terms of their business/production processes, products or services using digital technologies. They are based upon technology infrastructure and provide access to the latest knowledge, expertise and technology to support their customers with piloting, testing and experimenting with digital innovations. DIHs also provide business and financing support to implement these innovations, if needed across the value chain. A DIH is a regional multi-partner cooperation (including organisations like RTOs, universities, industry associations, chambers of commerce, incubator/accelerators, regional development
agencies and even governments) and can also have strong linkages with service providers outside of their region supporting companies with access to their services.

**Venture capital** Innovative and growth-oriented small businesses need to raise capital (equity investment) from external sources because they do not have their own resources or cannot access loans. Firms typically look for venture capital to provide them with the financing they need to expand, break into new markets, and grow faster. Although venture capital is only relevant for a small group of firms, it is essential for the growth of innovative firms. Venture capital funds raise a large part of their funding from institutional investors and they usually invest large amounts into firms with the potential for rapid growth.

**Forecast Scenarios**

In our methodology, the scenarios are used to elaborate the potential alternative growth paths of the European Data Economy, taking into account the main economic, technological, policy-regulatory and social factors affecting its development. The qualitative scenarios interact with the quantitative forecast models with a mutual fine-tuning and validation effect, by investigating the rationale behind potential growth trends, and vice-versa by taking into account insights from the data. The ultimate objective of the scenarios, however, is to analyse which combination of framework conditions and policy actions may maximise the growth potential of the European Data Market and Economy, and by feeding into the models estimate the actual size and depth of the potential benefits. In this way the scenarios provide a realistic approach to the forecast estimates – since we project a range of values (not a single estimate which may be widely off the mark) - and provide guidance on the potential consequences of different external events or alternative policy choices.

We have implemented the same scenario methodology used in the previous EDM Study. The scenario model used in this study is based on the definition of alternative assumptions about four main groups of key factors. IDC has developed and implemented this model in several projects about various ICT markets, from the Future Internet to Cloud Computing and the IoT and we believe it is thoroughly validated.

Every year the assumptions within each of the main groups of factors will be revised and updated or validated, or new ones will be added, leveraging the results of the research as well as IDC's periodically updated global Market Forecast Assumptions. These assumptions are collected, aggregated and shared with all IDC analysts at a global level by IDC's Global Research Group which is composed of experienced analysts and economists.

The selection of the most relevant factors in the scenario model was based on two main criteria:

- High level of impact on the development of the Data Market;
- High level of uncertainty, with potential different outcomes (assumptions) over the next five years.

The four main groups of factors are:

- Macroeconomic factors;
- Policy/regulatory conditions;
- Data Market dynamics factors;
- Global megatrends affecting all technology markets.

Even though they may seem obvious, still these four clusters correspond to the main typologies of factors which affect the evolution of the Data Market. Each cluster aggregates a set of interrelated
key factors; their combination differentiates the three scenarios (Figure 16). The scenarios are characterised by the interaction and co-dependency of these factors; no scenario can be explained only by one factor or one group of factors, not even GDP growth.

Figure 19: Structure of the Scenarios Model

Source: European Data Market Monitoring Tool, IDC 2015

The scenario model and the forecast indicators models are correlated.

The table below summarises the rationale of their selection and how their assumptions were used as inputs to the indicators' forecast models.

Table 17: Identification of main Factors driving the Scenarios

<table>
<thead>
<tr>
<th>Key Factors</th>
<th>Rationale</th>
<th>Inputs to the Forecast Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy/Regulatory conditions</td>
<td>Strong influence of the policy/regulatory framework on the model of development of the Data Market</td>
<td>Alternative policy and regulatory conditions by scenario</td>
</tr>
<tr>
<td>ICT Market/ FIWARE dynamics factors</td>
<td>Strong influence of alternative supply-demand dynamics on the market development paths</td>
<td>Alternative supply and take-up models by scenario</td>
</tr>
<tr>
<td>Global megatrends</td>
<td>Strong influence of global digital innovation trends on the EU Data Market growth</td>
<td>Alternative assumptions on the development of IoT, Cloud Computing, Mobile technologies based on IDC's 2025 forecasts</td>
</tr>
</tbody>
</table>

Measuring Data Professionals

Definition and Scope

Data professionals are workers who collect, store, manage, and/or analyse, interpret, and visualise data as their primary or as a relevant part of their activity. Data professionals must be proficient with the use of structured and unstructured data, should be able to work with a huge amount of data and be familiar with emerging database technologies.
In our definition, data professionals are not only data technicians but also data users who, based on more or less sophisticated tools, take decisions about their business or activity, after having analysed and interpreted available data. According to our definition, data professionals belong to the category of knowledge workers and specifically “codified” knowledge workers (Lundavall and Johnson, 1994); data professionals specifically deal with data while knowledge workers deal with information and knowledge.

The indicator has been measured according to the segmentations presented in the following table, including two sub-indicators about the share on employment and the intensity of employment.

Table 18: Indicator 1 – Data Professionals

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Number of data professionals</td>
<td>Total number of data professionals in the EU</td>
<td>Number, 2016-17-20 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: 28 EU MS + total EU By Industry: 11 industry sectors NACE rev.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Employment share</td>
<td>Total number as a share of total employment in the EU</td>
<td>% of total employment, 2017-18-19</td>
<td>By Geography: 28 EU MS + total EU By Industry: 11 industry sectors NACE rev.2 By Size: not applicable</td>
</tr>
<tr>
<td>1.3</td>
<td>Intensity share</td>
<td>Average number of data professionals per company (only for private sector)</td>
<td>Number, 2017-18-19</td>
<td>By Geography: 28 EU MS + total EU By Industry: 11 industry sectors NACE rev.2 By Size: not applicable</td>
</tr>
</tbody>
</table>

**Methodology Approach**

Our approach is based on an iterative process and on a calibration process of the final estimates. The approach has been repeated in the new study based on updates of the main sources.

**Statistical Identification**

Data professionals are not classified as such into any of the labour and occupation statistics. In order to define them statistically, we have adopted the International Standard Classification of Occupations (ISCO-08), selecting categories where data professionals may be included. The criteria adopted for the selection of the ISCO-08 codes are the following:

- We have selected the occupations where data professionals can be involved either as data providers or as data users;
- We have selected the occupations from 1 to 4-digit disaggregation;
- The occupation codes selected are those where the presence of data professionals can be detected because:
  - They hold deep analytical skills;
They do not need deep analytical skills but basics understanding of statistics and/or machine learning in order to conceptualise the questions that can be addressed through deep analytical skills;

- They are the ones providing enabling technology and therefore they are providers of data services.

- The selected codes are those where a significant part of the workers may be data professionals; the occupations where the data professionals are a very marginal part of the workers have been excluded; as an example, the medical practitioners have been excluded, although some practitioners may be data professionals because they undertake research activities. Since they are only a very marginal part of the practitioners, we excluded them from the occupations where data professionals are present;

- We excluded all the data professionals which are not included into the knowledge economy perimeter because their occupation is a low skilled one, i.e. with high routine level (as an example, call centre workers are in theory data professionals but since their activity is a routine one and as such excluded from the knowledge economy, they are not considered data professionals).

Table 19: ISCO-08 Structure and Data Professionals

<table>
<thead>
<tr>
<th>ISCO-08 structured Classification</th>
<th>Major Groups (1 digit)</th>
<th>Sub-groups (2 digits)</th>
<th>Minor Groups (3 digits)</th>
<th>Units (4 digits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of codes ISCO-08 structure</td>
<td>10</td>
<td>43</td>
<td>130</td>
<td>436</td>
</tr>
<tr>
<td>Number of selected codes including data professionals</td>
<td>4</td>
<td>9</td>
<td>21</td>
<td>52</td>
</tr>
<tr>
<td>Share of data professionals’ codes in the ISCO-08 structure</td>
<td>40%</td>
<td>21%</td>
<td>16%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Source: IDC elaboration on ISCO codes

**Calculation of the quantitative Perimeter**

The quantitative perimeter of employment where data professionals are trackable is based on the selected ISCO codes crossed with the NACE classification of economic activities, for each one of the 28 Member States and the EU as a whole, and has been updated based on the sources updates.

**Estimate and Calibration of the Penetration of Data Professionals**

The next step is the estimate of percentage of data professionals within the perimeter of data professional candidates. To calculate the coefficients for the calculation of such %, we have elaborated a set of assumptions (specified in the D2- Methodology report of the EDM Study). The assumptions have been revised and updated for each release of the study and applied to the model to calculate the share of data professionals by Member State and by industry.

**Forecasting Data Professionals**

The same model was applied to forecast data professionals to 2025, by developing specific assumptions by scenario, even though the level of uncertainty is higher, and the reliability of the forecasts is lower.
Measuring Data Companies

**Definition and Scope**

Data companies are organisations that are directly involved in the production, delivery and/or usage of data in the form of digital products, services and technologies. They can be both data suppliers’ and data users’ organisations:

- **Data suppliers** have as their main activity the production and delivery of digital data-related products, services, and technologies. They represent the supply side of the Data Market.
- **Data users** are organisations that generate, exploit collect and analyse digital data intensively and use what they learn to improve their business. They represent the demand side of the Data Market.

Table 20: Indicator 2 – Number of Data Companies

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Number of data suppliers</td>
<td>Total number of data suppliers, measured as legal entities based in the EU</td>
<td>Number, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: 28 EU MS + total EU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Industry: 2 NACE rev2 Section J Information and Communication and section M Professional, scientific and technical activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Company Size: below 250 employees above 250 employees</td>
</tr>
<tr>
<td>2.2</td>
<td>Share of data suppliers</td>
<td>Total data companies on total companies in industry J and M</td>
<td>% 2017-18-19</td>
<td>By Geography: 28 EU MS + total EU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Industry: 2 NACE rev2 Section J Information and Communication and section M Professional, scientific and technical activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Company Size: below 250 employees above 250 employees</td>
</tr>
<tr>
<td>2.3</td>
<td>Number of data users</td>
<td>Total number of data users in the EU, measured as legal entities based in one EU country</td>
<td>Number, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: 28 EU MS + total EU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Industry: 11 industry sectors NACE rev.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Company Size: below 250 employees above 250 employees</td>
</tr>
<tr>
<td>2.4</td>
<td>Share of data users</td>
<td>Total data users as share of total private companies</td>
<td>% 2017-18-19</td>
<td>By Industry: 11 industry sectors NACE rev.2</td>
</tr>
</tbody>
</table>

**Methodology Approach**

Data companies have been measured by updating the same model used in the previous EDM Study (see Figure below) which leverages both IDC and public sources.
• Eurostat business demography statistics in the European Union, treating aspects such as the total number of active enterprises in the business economy, their birth rates, death rates, and the survival rate (last update: December 2014);
• Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: March 2016);
• IDC’s detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 and 2015;
• IDC Worldwide Black Book (Standard Edition), quarterly updates from the years 2014 through 2015. The Black Book represents IDC’s quarterly analysis of the status and projected growth of the worldwide ICT industry in 54 countries.
• IDC End-User IT Trends and Digital Transformation: IDC European Vertical Markets Survey 2015

Figure 20: Data Companies Model

Measuring the Revenues of Data Companies

**Definition and Scope**

*Data companies’ revenues* are the revenues generated by data suppliers for the products and services specified in our definition of the Data Market. The revenues correspond to the aggregated value of all the data-related products and services generated by Europe-based suppliers, including exports outside the EU.

*Table 21: Indicator 3 – Revenues of Data Companies*
<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Total revenues of data companies</td>
<td>Total revenues of the Data Suppliers calculated by Indicator 2</td>
<td>Billion €, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: 28 EU MS + total EU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Company Size: below 250 employees</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>above 250 employees</td>
</tr>
<tr>
<td>3.2</td>
<td>Share of data companies’ revenues</td>
<td>Total revenues of the Data Suppliers calculated by Indicator 2</td>
<td>% of revenues on total, 2017-18-19</td>
<td>By Geography: 28 EU MS + total EU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Methodology Approach**

The indicator has been measured applying the same model used in the previous EDM Study, which calculated the revenues by feeding on:

- Eurostat and IDC statistics on average IT vendors revenues by size and sector;
- The total number of data companies by country, industry and size class;
- The value of the Data Market by country and industry;
- The estimated share of exports-imports in the value of the Data Market.

**Measuring the Data Market**

**Definition and Scope**

The Data Market is the marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data. We define its value as the aggregate value of the demand of digital data without measuring the direct, indirect and induced impacts of data in the economy as a whole. The value of the Data Market includes imports (data products and services bought on the global digital market from suppliers not based in Europe) and excludes the exports of the European data companies.

*Table 22: Indicator 4 – Size of the Data Market*

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Value of the Data Market</td>
<td>Estimate of the overall value of the Data Market</td>
<td>Billion €, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: total EU, EU28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Industry: 11 industry sectors NACE rev.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By Size: not applicable</td>
</tr>
</tbody>
</table>

**Methodology Approach**

The Data Market indicator is being updated every year for the duration of the study. The model is based on the extraction of data from IDC databases concerning the components of hardware,
software and services spending which fall in the definition of the Data Market. The IDC data is already segmented by country and by industry, even though not all Member States are covered, and the industry classification is slightly different from the one proposed in this project. The respective shares for the software, hardware, and services market used to derive the Data Market are derived from IDC surveys covering Big Data, IT spending patterns and intentions in the European market, and a survey of data suppliers and data users in key Member States, together with analyst expertise and alignment with IDC’s European and worldwide forecasts for the business analytics and Big Data Market.

The model updates the Data Market value shares by Member State and by industry.

*Figure 21: Data Market Model*

---

**Measuring the Data Economy**

**Definition and Scope**

The Data Economy measures the overall impacts of the Data Market on the economy as a whole. It involves the generation, collection, storage, processing, distribution, analysis elaboration, delivery, and exploitation of data enabled by digital technologies. The Data Economy also includes the direct, indirect, and induced effects of the Data Market on the economy.

The Data Economy indicator measures the value of the Data Economy based on the estimate of all the economic impacts following the adoption of data-driven innovation and data technologies in the EU. As such, the indicator aggregates direct, indirect, induced impacts of the Data Market defined as follows.

1. **The direct impacts**: these are impacts generated by the data industry itself; they represent the activity engendered by all businesses active in the data production. The quantitative direct impacts are measured by the revenues from data products and services sold, i.e. the value of...
the Data Market. We prefer to adopt the Data Market value as a good proxy of the direct impacts because its estimates are more reliable than the value of the revenues. **The direct impacts**: the initial and immediate effects generated by the data suppliers; they represent the activity potentially engendered by all businesses active in the data production. **The quantitative direct impacts have then been measured as the revenues from data products and services sold, i.e. the value of the Data Market.** As Data Market estimation is more reliable than data companies’ revenues estimation, we consider the Data Market value as a good proxy of the direct impacts. Therefore, for the sake of simplicity, direct impacts do coincide with the value of the Data Market.

2. **The indirect impacts**: the economic activities generated along the company's supply chain by the data suppliers. There are two different types of indirect impacts: the backward indirect impacts and the forward indirect impacts (Richardson, 1985):
   a. **The backward indirect impacts**: such impacts represent the business growth resulting from changes in sales from suppliers to the data industry. In order to produce and deliver data products and services, the data companies need inputs from other stakeholders. Revenues from those sales to data companies are the backward indirect impacts.
   b. **The forward indirect impacts**: such impacts include the economic growth generated through the use of data products and services by the downstream industries, i.e. the data users as a selected number of industries. For the user companies, data is now a relevant factor of production; the adoption of data products and services by the downstream industries provides different types of competitive advantage and productivity gains to the user industries. The main benefits that the exploitation of data can provide to downstream industries are (OECD, 2013, Mc Kinsey, 2011):
      i. Optimising production and delivery processes: data-driven processes (data-driven production);
      ii. Improving marketing by providing targeted advertisements and personalised marketing practices (data-driven marketing);
      iii. Improving existing organisation and management practices (data-driven organisation).

3. **The induced impacts**: these impacts include the economic activity generated in the whole economy as a secondary effect. Induced additional spending is generated both by new workers, who receive a new wage, and by the increased wage of existing jobs. This spending induces new revenues creation in nearly all sectors of the economy. The additional consumption does support economic activity in various industries such as retail, consumer goods, banks, entertainment, etc.

*Table 23: Indicator 5 – Value of the Data Economy*

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Value of the Data Economy</td>
<td>Value of the Data Market plus direct, indirect and induced impacts on the EU economy</td>
<td>Billion €, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: Total EU + EU 28</td>
</tr>
<tr>
<td>5.1</td>
<td>Incidence of the Data Economy on GDP</td>
<td>Ratio between value of the Data Economy and EU GDP</td>
<td>%, 2017-18-19 Forecast to 2025, 3 Scenarios</td>
<td>By Geography: EU 28 + Total EU</td>
</tr>
</tbody>
</table>
This estimate of the Data Economy does not include the user benefits and social impacts of data-driven innovation such as changes in quality of life (health, safety, recreation, air quality). Although these benefits may be evaluated in economic (money) terms, they are not economic impacts as such and as defined above as they do not induce an increase in the business activities and a consequent growth in GDP.

Analysts underlined that the new decision-making processes act as a rationalisation and optimisation factor (Brynjolfsson, 2011, Mc Kinsey, 2012), since they improve effectiveness and efficiency, and in some cases, they may have a disruptive effect. The impacts related to the new decision-making processes are the one we have called the forward indirect impacts.

The value creation process based on data rests on the elaboration of information and knowledge (OECD 2016), although the boundaries between data, information, and knowledge are sometimes fuzzy. The huge volume of data is a global phenomenon which is sometimes viewed with suspicion by citizens, consumers and businesses because data flows are seen as an intrusion of the privacy. Nevertheless, there is currently some evidence showing that data analysis can provide benefits to both businesses and consumers. By the way, this is not surprising since we should remind that the economic theory holds that information encourages competition between businesses for the benefit of consumers.

Data do not provide value and benefits as such; data need to be collected, stored, aggregated, combined and analysed in order to be appropriately used for decision making processes. To create value, data need to be processed (OECD, 2016):

- **Extracting information from structured and unstructured data:** data analytics techniques are today able to analyse both structured and unstructured data. We should remind here that most data stored by businesses are unstructured (IDC, 2012). Technologies such as optical character recognition, natural language processing, face recognition algorithms and machine learning algorithms are empowering the use of all data.

- **Real-time monitoring and tracking:** analysis of data in real time is often mentioned as one of the most powerful factor since it supports organisations to make real-time decisions, which, in a fast-changing world, is a well-known competitive advantage.

- **Inference and prediction:** until now, prediction was based exclusively on prior information and data series. Data analytics can now enable the creation of information even without prior information. Such information can be created through patterns and correlations of data. Personal information, for example, can be deduced from anonymous or non-personal data. Businesses and organisations demand real time insights rather than historical and periodical information, and for advanced specialised data analytics services. Algorithms allow machine and statistical learning based on non-specific data; businesses can learn and predict a lot about their customers even if they do not have specific data and time series about the issue they are interested in. Machine learning has, as an example, applications in health care where data collected on patients are recorded by imaging, or it supports production processes to increase the quality of production.

The diffusion of technology supporting production and analysis of data induces organisations and businesses to base their decisions on data much more than they were used to do. As pointed out by OECD in its recent report, the process to take decisions is also changing. Decision makers do not necessarily need to understand the phenomenon before they act on it. A store can change the product placement based on data analysis without the need to know the reason why such a change should improve the sales. There is therefore a decision automation process: “first comes the analytical factor, then the action, and last, if at all, the understanding” (OECD, 2015).
The impacts of such a new approach to decision making and to the use of data in all the enterprises and organisations’ functions are many and varied, so that we believe, such impacts will be object of studies and analysis in the upcoming years. It is, at this point, difficult to classify them and to suggest a taxonomy of such impacts.

Such impacts have been observed through some empirical studies and case analysis. The most relevant ways the benefits appear are the following.

- **Creating more information, knowledge and transparency**: technology is making data more accessible and exploitable to all kind of stakeholders, including SMEs. This increases transparency and decisions are made on a rational process.

- **Improving performance**: having access to a wide information and to a high number of data is changing the way of making decisions. An increasing number of organisations are going to become data-driven organisations, which means that they make decisions based on empirical results. As an example, retailers can adjust prices and promotions, more precisely than they were used to and in real time. This may improve competitiveness. McKinsey underlines that the health sector is achieving a lot of benefits from the new making decisions process: studies on clinical data allow to identify and understand the sources of variability in treatment, to identify the best treatment protocols and to create guidelines for the optimization of treatment decisions. This does not only increase the effectiveness of treatments, but it also produces saves.

- **Improving customization of actions for better decisions**: data technology is definitely improving the segmentation of customers and the analysis of their preferences in real time. This allows companies to supply products and services targeted to specific groups of individuals who have specific needs and preferences. Such a segmentation is also useful when supplying public services. Such a segmentation helps define the price precisely and offering exactly what is needed which means a better quality and also companies avoid offering products and services the consumers are not willing to pay.

- **Innovating products and services as well as business models**: the more information and understanding businesses have about their customers, the better they can serve them. It is important to say that although consumers may fear their privacy is injured, this can also provide them unexpected surplus: real time price comparison services do not only provide better transparency but also allow buying the best product at the most convenient price (for example when buying online airline tickets or when booking hotels). Companies can in fact produce and create new products and services to better satisfy their customers’ needs. This is true also for the public sector and specifically for the health care system where preventing care programs can be created.

These effects are reflected in an increase in revenues due to higher market share from the increase in competitiveness or due to a reduction in costs. All these effects are included in the forward indirect impacts; these impacts are delivered on the user industry, and because of the above reasons, these are the impacts we consider new on the overall economic system.

**Methodology Approach**

Measuring the Data Economy depends on the macroeconomic context on one hand, and on the adoption/diffusion and integration processes the companies are implementing on the other hand. Moreover, there is a necessary time lag before the impacts take place in the economic system. Therefore, the estimates are based on a set of assumptions, including choices about proxy indicators.

In order to measure the impact of the diffusion and use of data services and products, we estimated each component (as defined in the above paragraph) of the impact separately.
The estimation approach developed in the previous study was based on a number of assumptions on one hand and on results from a survey launched during the first-year research.

The following assumptions have been confirmed:

- The penetration rates of data in terms of value added for the user industries using data are positively correlated to the penetration rate in terms of number of companies using data.
- The survey conducted in the study 2013-2016 provided information about the quantitative benefits due to the use of data, for the six major Member States plus Czech Republic; such benefits have been taken into consideration for the six major Member States.
- For Austria, Belgium, Denmark, Finland, Ireland, Luxembourg, Malta, the Netherlands, and Sweden we assumed that these Member States have the same distribution of benefits as the average of the Big Six.
- For the other Member States, we estimated the benefits of the rest of Europe, based on the survey results, and we assumed that all the minor Member States are achieving benefits similar to the rest of Europe.
- For the induced impacts, we assumed that the additional earnings are spent according to the general economic mood.

In order to update the estimates of the different components of the impacts, we have adopted some new assumptions:

- In the next three years, we are going to stay in a relatively emerging stage of the data diffusion, so that in our view the structure of the data impacts is not going to change.
- For the quantitative benefits due to the use of data, we assume that the benefits will quantitatively vary and be correlated to the macroeconomics trends and specifically with the industries’ trends (and stakeholders) affected.

**Measuring the Data Skills Gap**

**Definition and Scope**

This indicator captures the potential gap between demand and supply of data skills in Europe, since the lack of skills may become a barrier to the development of the data industry and the rapid adoption of data-driven innovation.

**Table 24: Indicator 6 – Data Skills Gap**

<table>
<thead>
<tr>
<th>N.</th>
<th>Name</th>
<th>Description</th>
<th>Type and Time</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Data Workers Skills Gap</td>
<td>Gap between demand and supply of data workers</td>
<td>Absolute number and % on total demand, 2017-18-19 Forecast to 2025, 3 scenarios</td>
<td>By Geography: total EU28; main EU Member States</td>
</tr>
</tbody>
</table>

**Methodology Approach**

The methodology approach is the same implemented by IDC-empirica to estimate the supply-demand balance of ICT skills in the EU (e-Skills) on behalf of the EC DG Enterprise (now DG GROW). The model was first developed in 2009 and since then has been successfully validated and updated.
several times. The results have been used by the EC to support the e-skills policy and the latest results were presented in December 2015 at the European E-skills 2015 Conference in Brussels14. However, data skills are not a subset of ICT skills so the scope of supply and the dynamics of demand are different from the e-skills model developed by IDC.

To update the measurement of the indicators the study team has applied the same model developed for the previous EDM Study, combining the estimates and forecasts of the demand and supply of data professionals with data skills leveraging a wealth of different sources, among which:

- European Data Science Academy (EDSA) project deliverables and publications (July 2015).

Figure 22: The Data Skills Demand-Supply Balance Model

![Diagram of Data Skills Demand-Supply Balance Model](image)

Source: European Data Market Monitoring Tool, IDC 2016

---

ESSENTIAL GLOSSARY – THE KEY INDICATORS

Data professionals are data workers who collect, store, manage, and/or analyse, interpret, and visualise data as their primary or as a relevant part of their activity. Data professionals must be proficient with the use of structured and unstructured data, should be able to work with a huge amount of data and be familiar with emerging database technologies. They elaborate and visualise structured and unstructured data to support analysis and decision-making processes.

Data companies can be both data suppliers’ and data users’ organisations:

- **Data suppliers** have as their main activity the production and delivery of digital data-related products, services, and technologies. They represent the supply side of the Data Market.
- **Data users** are organisations that generate, exploit, collect and analyse digital data intensively and use what they learn to improve their business. They represent the demand side of the Data Market.

Data companies’ revenues are the revenues generated by data suppliers for the products and services specified in our definition of the Data Market. The revenues correspond to the aggregated value of all the data-related products and services generated by Europe-based suppliers, including exports outside the EU.

The Data Market is the marketplace where digital data is exchanged as “products” or “services” as a result of the elaboration of raw data. We define its value as the aggregate value of the demand of digital data without measuring the direct, indirect and induced impacts of data in the economy. The value of the Data Market includes imports (data products and services bought on the global digital market from suppliers not based in Europe) and excludes the exports of the European data companies.

The Data Economy measures the overall impacts of the Data Market on the economy. It involves the generation, collection, storage, processing, distribution, analysis elaboration, delivery, and exploitation of data enabled by digital technologies. The Data Economy also includes the direct, indirect, and induced effects of the Data Market on the economy.

The Data Professionals’ Skills Gap captures the potential gap between demand and supply of data skills in Europe, since the lack of skills may become a barrier to the development of the data industry and the rapid adoption of data-driven innovation.

Data is usually defined as qualitative or quantitative statements or information which can be coded and which are assumed to be factual and not the product of analysis or interpretation. For the sake of this study we consider only data which is collected, processed, stored, and transmitted over digital information infrastructures and/or elaborated with digital technologies. This definition includes multimedia objects which are collected, stored, processed, elaborated and delivered for exploitation through digital technologies (for example, images databases).

Information is the output of processes that summarise, interpret or otherwise represent the content of a message to convey meaning. Therefore, information is not a mere synonymous of data.

The Knowledge Economy is defined as the production of products and services based on knowledge-intensive activities that contribute to an accelerated pace of technical and scientific advance, as well
as rapid obsolescence. The key component of a knowledge economy is a greater reliance on intellectual capabilities than on physical inputs or natural resources.

The Internet Economy is defined as covering the full range of our economic, social and cultural activities supported by the Internet and related information and communications technologies\(^\text{15}\).

Information or Knowledge workers in the most basic definition are persons employed to produce or analyse ideas and information. Multiple sources define knowledge workers as workers creating knowledge capital, who process existing information to create new information to be used to define and solve problems. They include, as an example, medical practitioners, lawyers, judges, teachers, architects, engineers, managers or salespeople. Their main capital is knowledge, and they are mainly focused on “non-routine” tasks.

Data workers collect, storage, manage and analyse data, as their primary activity. Data workers can be knowledge workers if they are focused on non-routine tasks. For example, data entry clerks’ primary activity is related to data, so they are data workers. However, data entry is a very routine task and as such data entry clerks should not be considered as knowledge workers. Another category of data workers is data analysts, who usually extract and analyse information from one single source, such as a CRM database. They require a medium level of creative thinking and usually work on structured data.

Data scientists require solid knowledge in statistical foundations and advanced data analysis methods combined with a thorough understanding of scalable data management, with the associated technical and implementation aspects (European Big Data Value Partnership Strategic Research and Innovation Agenda, April 2014). They can deliver novel algorithms and approaches such as advanced learning algorithms, predictive analytics mechanisms, etc. Data scientists should also have a deep knowledge of their businesses; the most difficult skills to find, include advanced analytics and predictive analysis skills, complex event processing skills, rule management skills, business intelligence tools, data integration skills (UNC, 2013).