IDC
Analyze the Future

**European Data Market
SMART 2013/0063**

Final Report

IDC
Analyze the Future

**OPEN
EVIDENCE**

February 1st 2017

This study was prepared for the European Commission (Directorate-General for Communications Networks, Content and Technology) by IDC and Open Evidence.

DISCLAIMER

SMART 2013/0063 European Data Market

Table of Content

SMART 2013/0063 European Data Market

ABSTRACT

This report presents a set of indicators measuring the European population of data workers, the value of the data market, the number of data user enterprises, the number of data companies and their revenues, and the overall value of the impact of the data economy on EU GDP. All indicators are presented for the years 2013 through 2016 and forecasted to 2020, exploring three alternative potential scenarios of evolution: Baseline, High Growth and Challenge scenarios.

In particular:

- The total number of **data workers**, their share on the total employment in the EU and their intensity (i.e.: their average number per company) have constantly increased throughout the period under consideration;
- **Data companies** - the organisations providing data (data-suppliers) and making a strong reliance on data (data-users) - have increased in number and share on the total number of companies in the EU from 2013 to 2016 and are projected to continue their growth throughout 2020 under all three forecast scenarios;
- The value of the overall **data market** (i.e. the market where digital data is exchanged as products or services derived from raw data) and the value of the overall **data economy** (including the economic impacts generated by data-suppliers, data users and the whole) present the most dynamic picture and are expected to further increase up to 2020 under the three scenarios;
- The **data worker skill gap** - the potential gap emerging between the demand and supply of data workers - reveals a potential lack of supply of data skill in Europe across the period under consideration, with specific reference to the High-Growth scenario;
- Finally, the report complements its "business orientation" by looking at the potential benefits that the citizens as a whole can obtain from the data economy in Europe. In this context, the **citizen's reliance on the data market** is measured through an ad-hoc indicator which shows a positive trend throughout the period 2013-2016.

1 INTRODUCTION

1.1. Overview

This document represents the Final Report of the European Data Market SMART 2013/0063 study entrusted to IDC and Open Evidence by the European Commission, DG Connect.

This document focuses on the following set of indicators:

- Indicator 1.1 Number of data workers
- Indicator 1.2 Employment share of data workers
- Indicator 1.3 Intensity share of data workers
- Indicator 2.1: Number of data companies (suppliers)
- Indicator 2.2: Share of data companies (suppliers)
- Indicator 2.3: Number of data users
- Indicator 2.4: Share of data users
- Indicator 3.1: Revenues of data companies
- Indicator 3.2: Share of data companies' revenues
- Indicator 4.1: Value of the data market
- Indicator 4.2: Value of the data economy
- Indicator 5.1: Data workers' skills gap
- Indicator 6.1: Citizens' data market.

Each indicator is measured at the level of the total EU28 and for all 28 EU Member States, when available and applicable; industry-specific and company-size views are also offered with indicators provided by industry sector and company size bands, when possible.

Two different views of the data are presented at European level: the usual EU28, and the EU27 which exclude the United Kingdom and refers to the UK's vote to leave the EU (Brexit).

The table below offers an overview of the full set of indicators that have been developed in this Final Report.

Table 1 Indicators delivered in the Final Report

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
1.1	Number of data workers	2013	Delivered	Delivered	Not applicable	Delivered	Delivered
		2014	Delivered	Delivered	Not applicable	Delivered	Delivered
		2015	Delivered	Delivered	Not applicable	Delivered	Delivered
		2016	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Challenge	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 High Growth	Delivered	Delivered	Not applicable	Delivered	Delivered
1.2	Employment Share	2013	Delivered	Delivered	Not applicable	Delivered	Delivered
		2014	Delivered	Delivered	Not applicable	Delivered	Delivered
		2015	Delivered	Delivered	Not applicable	Delivered	Delivered
		2016	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 Challenge	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 High Growth	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
1.3.	Intensity Share	2013	Delivered	Delivered	Not applicable	Delivered	Delivered
		2014	Delivered	Delivered	Not applicable	Delivered	Delivered
		2015	Delivered	Delivered	Not applicable	Delivered	Delivered

SMART 2013/0063 European Data Market

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
		2016	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 Challenge	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 High Growth	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
2.1	Number of Data Companies (suppliers)	2013	Delivered	Delivered	Delivered	Delivered	Delivered
		2014	Delivered	Delivered	Delivered	Delivered	Delivered
		2015	Delivered	Delivered	Delivered	Delivered	Delivered
		2016	Delivered	Delivered	Delivered	Delivered	Delivered
		2020 Baseline	Delivered	Delivered	Delivered	Delivered	Delivered
		2020 Challenge	Delivered	Delivered	Delivered	Delivered	Delivered
		2020 High Growth	Delivered	Delivered	Delivered	Delivered	Delivered
2.2	Share of Data Companies (suppliers)	2013	Delivered	Delivered	Not applicable	Delivered	Delivered
		2014	Delivered	Delivered	Not applicable	Delivered	Delivered
		2015	Delivered	Delivered	Not applicable	Delivered	Delivered
		2016	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 Challenge	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 High Growth	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
2.3	Number of	2013	Delivered	Delivered	Delivered	Delivered	Delivered

SMART 2013/0063 European Data Market

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
	Data Users	2014	Delivered	Delivered	Delivered	Delivered	Delivered
		2015	Delivered	Delivered	Delivered	Delivered	Delivered
		2016	Delivered	Delivered	Delivered	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered	Delivered	Delivered	Delivered
2.4	Share of Data Users	2013	Delivered	Delivered	Delivered	Delivered	Delivered
		2014	Delivered	Delivered	Delivered	Delivered	Delivered
		2015	Delivered	Delivered	Delivered	Delivered	Delivered
		2016	Delivered	Delivered	Delivered	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered	Not applicable	Delivered	Delivered
3.1	Revenues of Data Companies	2013	Not applicable	Delivered	Delivered	Delivered	Delivered
		2014	Not applicable	Delivered	Delivered	Delivered	Delivered
		2015	Not applicable	Delivered	Delivered	Delivered	Delivered
		2016	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered	Delivered	Delivered	Delivered

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
3.2	Share of Data Companies' Revenues	2013	Not applicable	Delivered	Delivered	Delivered	Delivered
		2014	Not applicable	Delivered	Delivered	Delivered	Delivered
		2015	Not applicable	Delivered	Delivered	Delivered	Delivered
		2016	Not applicable	Delivered	Delivered	Delivered	Delivered
		2020 Baseline	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 Challenge	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
		2020 High Growth	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
4.1	Value of the Data Market	2013	Delivered	Delivered	Not applicable	Delivered	Delivered
		2014	Delivered	Delivered	Not applicable	Delivered	Delivered
		2015	Delivered	Delivered	Not applicable	Delivered	Delivered
		2016	Delivered	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered	Not applicable	Delivered	Delivered
4.2	Value of the Data Economy	2013	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2014	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2015	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2016	Not applicable	Delivered	Not applicable	Delivered	Delivered

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
		2020 Baseline	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered	Not applicable	Delivered	Delivered
4.3	Incidence of the Data Economy on GDP	2013	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2014	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2015	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2016	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
5.1	Data Worker skills gap	2013	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2014	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2015	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2016	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
		2020 Challenge	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
		2020 High Growth	Not applicable	Delivered (Big Six)	Not applicable	Delivered	Delivered
6.1	Citizens'	2013	Not	Delivered	Not	Delivered	Delivered

#	Name of Indicator	Year	Industry	Member State	Company Size	EU28 Level	EU27 Level
	Data Market		applicable		applicable		
		2014	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2015	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2016	Not applicable	Delivered	Not applicable	Delivered	Delivered
		2020 Baseline	Not applicable	Not applicable	Not applicable	Not applicable	Delivered
		2020 Challenge	Not applicable	Not applicable	Not applicable	Not applicable	Delivered
		2020 High Growth	Not applicable	Not applicable	Not applicable	Not applicable	Delivered

To enable a meaningful comparison of the European data market and data economy within the main EU partners worldwide, the study team has also produced a set of indicators for Brazil, Japan and the United States. In the Final Report we have updated the values and results for the following indicators:

- Indicator 1.1 Number of data workers
- Indicator 1.2 Employment share
- Indicator 2.1 Number of data companies
- Indicator 3.1 Revenues of data companies
- Indicator 4.1 Value of the data market
- Indicator 4.2 Value of the data economy
- Indicator 4.3 Incidence of the data economy on GDP

These indicators are presented in Chapter 12.

1.2. Structure of the Document

This document is organized along the following chapters:

- Chapters 1 and 2 include a brief introduction and a short reminder of the overall study's goals and objectives as well as a summary of the European Data Market Monitoring Tool and its functioning.
- Chapter 3 covers the three evolution scenarios of the data market and the methodological approach to the indicators' forecasts to 2020, with specific reference to the assumptions' changes that have been introduced to refresh the data of this Final Report.
- Chapter 4 is devoted to the latest measurement of the data workers including their main values in absolute terms, their share in terms of total employment and their forecast to 2020 according to the three scenarios under consideration.

Furthermore, the chapter proposes a more precise definition of data workers taking into account the existing definitions of similar concepts in the current literature.

- Chapters 5 and 6 provides an update of the indicators measuring the data companies (both suppliers and users of data) in terms of absolute numbers and produced revenues. It includes the updated forecast of the indicator by 2020.
- Chapter 7 presents the updated values of the indicators measuring the size of the data market in Europe based on the total spending on software, hardware, and IT services' technologies and its contribution to the data market in Europe. It provides the forecast of the indicator by 2020.
- Chapter 8, Measuring the Data Economy for the years 2014 and 2015, as well as the forecasts to 2020 according to the three scenarios under consideration.
- Chapter 9 summarizes the qualitative findings on the data market functioning and evolutions in the light of the most recent stories produced in the course of the study.
- Chapter 10 is devoted to the update of the data workers' skills gap indicator in the EU.
- Chapter 11 provides a new measurement of the citizens' reliance on the data market in Europe.
- Chapter 12, Worldwide Monitoring of the Data Market, will be released as soon as the new statistical sources will be made available.
- Chapter 13 presents the set of recommendations stemming leveraging the results obtained in the study.
- Chapter 14 provides a set of concluding remarks and a summary of the report.
- The Methodological Annex summarizes the key methodological steps that we have undertaken to measure the indicators covered in both the previous reports and in the current report.
- The Statistical Annex presents the detailed tables by Member State, by Industry, by share, including the forecast data by the three scenarios. The tables are ordered by indicator.

1.3.   Main Goals and Objectives of the Study

The main goal of the European data market study is to define, assess, and measure the European data economy, supporting the achievement of the Data Value Chain policy of the European Commission, which aims to develop a vibrant and innovative data ecosystem of stakeholders driving the growth of this innovative market in Europe.

The study is built around the accomplishment of three main objectives (the development of a European Data Market Monitoring Tool, the production of descriptive stories about the data economy, and the building of stakeholders' community), with the overarching goal to feed its results into the European Commission's Digital Agenda.

1.4.   Methodology Approach

The measurement of the indicators is based on a sophisticated methodology combining desk research, data collection, and the development of five different models for the main indicators (data workers, data suppliers, data users, data market, and data economy). The data collection included two ad hoc surveys carried out in February – March 2015 in eight EU Member States (Czech Republic, France, Germany, Italy, Spain, Poland, Sweden

and the U.K.). The two surveys targeted potential data companies in the ICT and professional services industries (235 interviews) and data users in eleven industries (1,184 interviews). The data companies' survey was smaller because it targeted only two sectors. The models leveraged IDC's proprietary databases and forecasts to 2020 for the main IT markets.

The model forecasts were based on the estimates of key macroeconomic indicators (EU GDP, EU total ICT spending, and unemployment) and the assumptions for the three scenarios, as well as IDC's current forecasts to 2020.

This final round of measurement of the European Data Market Monitoring Tool has been carried out following the main steps of the methodological approach, which was adopted to produce the indicators and the results of the Final Report as specified in the Methodology Annex of this document (see: Chapter 15: Methodology Annex). However, in order to obtain the most up-to-date figures for each of the indicators featured in in the European Data Market Monitoring Tool, and in the absence of fresh primary research data collected through a new, ad-hoc survey, the study team has undertaken additional desk research, leveraging both internal IDC sources, as well as public sources from EU and national statistical offices. In particular:

- IDC updated the indicators on data market, data companies, data companies' revenues, and the data economy by leveraging a variety of inputs, including but not limited to:
  - o 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: November 2016);
  - o Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: June 2016);
  - o IDC's detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014,2015, and 2016;
  - o IDC Worldwide Black Book (Standard Edition), quarterly updates form the years 2014, 2015 and 2016. The Black Book represents IDC's quarterly analysis of the status and projected growth of the worldwide ICT industry in 54 countries.
  - o IDC 2016 End-User IT Trends and Digital Transformation: IDC European Vertical Markets Survey 2015, (December 2015)
  - o IMF World Economic Outlook (WEO) Database, October 2016
  - o Consensus Forecasts, Consensus Economics, monthly updates (latest update: November 2016)
- The same sources were used to estimate the indicators on Data Workers and Data Workers' Skill Gap. For these two specific indicators, however, the study team also leveraged the following sources:
  - o OECD Digital Economy Papers, among which: OECD (2014), Measuring the Digital Economy: A New Perspective; OECD Digital Economy Outlook 2015
  - o ILOSTAT (International Labour Organization) Statistics and Databases (2016)
  - o EUROSTAT Tertiary Education Statistics (Last update: December 2015).
  - o European Data Science Academy (EDSA) project deliverables and publications (July 2016).
- IDC estimated the indicator on the Citizens' Reliance on the Data Market by using the above-mentioned sources plus the following sources:

- o The Digital Economy and Society Index (DESI), Human Capital Dimension, (2a Basic Skills and Usage; 2b Advanced skills and Development), last update: June 2016.
- o IDC European Quarterly Wearables Tracker Results: Western Europe 1Q16 Analysis, September 2016
- o IDC FutureScape: Worldwide Wearables and Augmented Reality/Virtual Reality 2017 Predictions, November 2016.

Of particular relevance among the above-mentioned sources were IDC annual SMB and Vertical Markets end user surveys /IDC End-User IT Trends and Digital Transformation: IDC European Vertical Markets Survey 2016) 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 2016 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 Workers, Data Workers' Skills Gap and Citizens' Reliance on the Data Market for the year 2015 and for the three 2020 scenarios.

A comprehensive and detailed description of this report's methodology approach is offered in the Methodology Annex at the end of the document.

1.5. Essential Glossary: Indicators' Definitions

1.5.1 Indicator 1 – Data Workers

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

Indicator 1.2 – Employment Share

Employment share is given by the share of data workers on the total employment in Europe, in percentage

Indicator 1.3 – Intensity Share

This indicator measures the average number of data workers calculated on the total number of data user companies.

1.5.2 Indicator 2 – Data Companies

Data companies are data suppliers' organizations, whose main activity is the production and delivery of digital data-related products, services, and technologies. They represent the supply side of the data market. On the other hand, 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.

Indicator 2.2 – Share of Data Companies

This indicator measures the share of data companies (defined above) on the total number of companies in the ICT and Professional services industries in Europe, expressed in percentage.

Indicator 2.3 and Indicator 2.4 – Number of data user companies, Share of data user companies

These indicators measure the number of European data users, counted as legal entities based in one EU Member State, and as a share of the total number of private enterprises in the EU.

1.5.3 Indicator 3.1 – Data Companies Revenues

This indicator measures the revenues of the data companies identified and classified by Indicator 2, 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 companies, including exports outside the EU.

Indicator 3.2 – Share of Data Companies Revenues

This indicator expresses the ratio between the data companies' revenues, defined above, and the total amount of companies revenues in sectors J and M

1.5.4 Indicator 4.1 Value of 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. 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 (please see indicator 4.2 "Value of the Data Economy"). The value of the data market is not exactly equal to the aggregated revenues of the European data companies because it 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.

Indicator 4.2 Data Economy

This indicator measures the value of the Data Market (defined above) plus the estimate of direct, indirect and induced impacts on the economy.

1.5.5 Indicator 5 – Data Workers Skills Gap

This indicator captures the potential gap between demand and supply of data worker skills in Europe. Demand and supply are estimated separately taking into account the supply of graduates, the level of unemployment, and the entry and exit flows of the data worker market. The main goal of this indicator is therefore to verify the potential existence of a lack of supply of data skills in Europe which may become a barrier to the development of the data industry and the rapid adoption of data-driven innovation.

1.5.6 Indicator 6 - Citizens' Reliance on the Data Market

This indicator measures the level of citizens' reliance on data. It aims to complement the "business orientation" of the other indicators by providing a snapshot of how citizens are taking advantage of data-driven solutions in their daily life.

2 MONITORING THE EUROPEAN DATA MARKET

The results presented in this Final Report follows the key methodological steps that the study team developed for this study, that is:

- The development of a detailed European Data Market Taxonomy
- The design of the Data Value Chain to reflect the data ecosystem in Europe
- The design of a European Data Market Monitoring Tool to be used for collection and measurement of facts and figures on the European data market size and trends.

A very brief account of these steps is provided below.

2.1. The European Data Market Taxonomy

The European data market taxonomy presents clear definitions of all the main terms used in the analysis and in the monitoring tool, providing an objective and scientific basis for the definitions of the indicators and the scope of their measurements.

The taxonomy has been developed on the basis of desk research on the main public sources and IDC's own taxonomies, research, and body of knowledge.

The taxonomy adopted in this report includes definitions used for:

- Data and type of data
- Data market, data economy, data workers, data scientists, data companies
- Data skills
- Data-based products and services
- Main stakeholders
- Main framework conditions

The complete taxonomy adopted in this report is included in the European Data Market Methodology Report (Deliverable D2), Annex 1.

2.2. The Data Value Chain

The taxonomy used in this report has been complemented by the design of the full Data Value Chain in use within the European data market.

The data value chain's components and the data ecosystem adopted by the study team is summarized in Figure 1 below.

Figure 1 The Data Value Chain and the data ecosystem
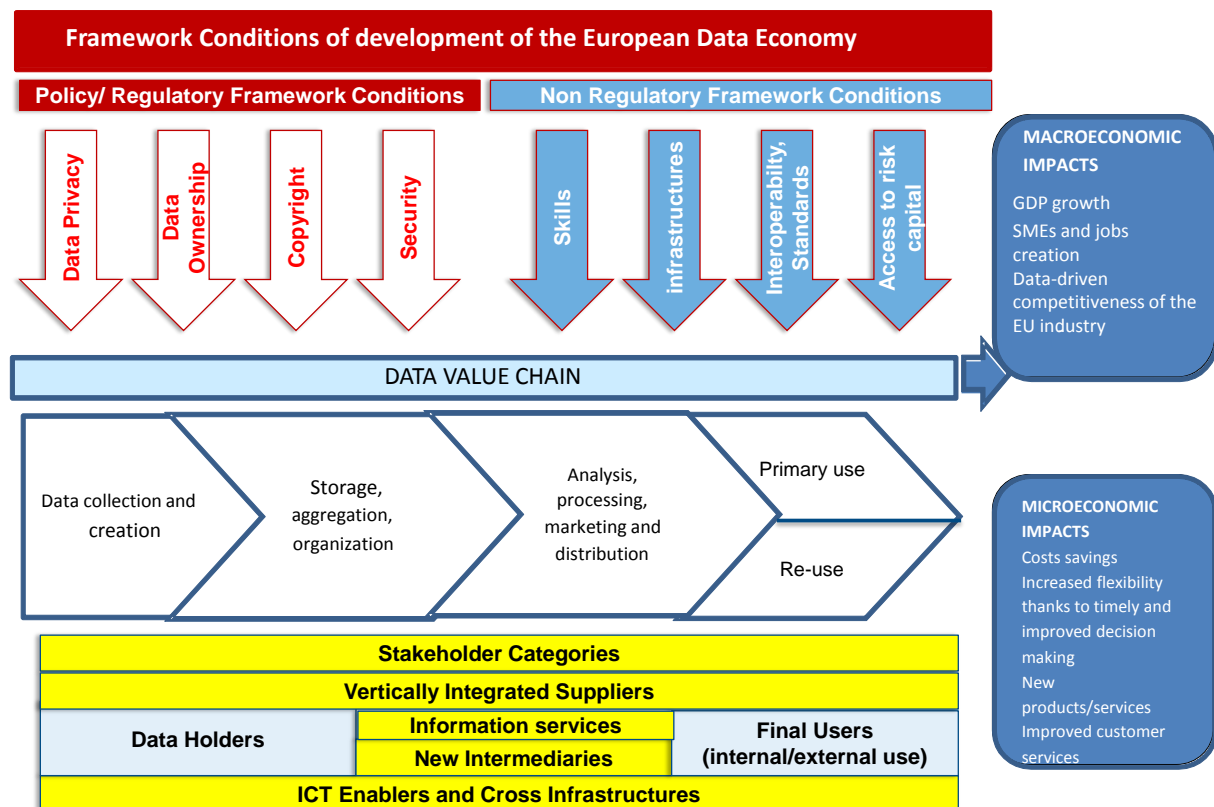
Source: IDC 2013

Figure 1 comprises the following elements that describe the structure of the data economy:

- The data value chain shows the four main phases of manipulation of data which lead to its exploitation.
- The macroeconomic and microeconomic impacts identify the direct and indirect impacts of the data value chain on the economic system and user enterprises.
- The stakeholder categories identify the main type of players on the basis of their role in the data value chain.
- The framework conditions identify the main factors which will enable or prevent the development of the European data market and economy. They are divided into policy-regulatory framework conditions and non-regulatory conditions.

Within the framework of the present study, the main steps in the data value chain to be taken into consideration are as follows:

- **Collection/access of data** from myriad sources within the applicable legal framework. Collection can be direct (for example, through loyalty schemes operated by retailers, transport, and hospitality service providers) or indirect (for example, by recording the location of someone using a cellular phone). Data can be also created through analysis rather than being captured.
- **Storage and aggregation** by service providers and social networks, but also by companies in traditional sectors such as finance, retail, transport, utilities, and government.

SMART 2013/0063 European Data Market

- **Analysis, processing, marketing, and distribution**, merging data from different sources (public, proprietary, or institutional research) and relying on analytics to derive insights and value. Traditional players across vertical markets can perform this task if they have the necessary skills/technology; alternatively, they can rely on external data brokers and providers.
- **Usage**, both in the public and private sectors, to better serve customers and/or improve efficiency. The use of data is broken down between primary use (when data is used for the goal for which it is collected: for example, mobile traffic data to bill customers by a telecom company) and secondary use or reuse (when data is exploited for other goals, for example when mobile traffic data is used to map customer movements for a retail company). Reuse is expected to be the source of much of the value added in the data market.

2.3. The European Data Market Monitoring Tool

The study team has further designed a European Data Market Monitoring Tool to be implemented during the activities of data collection and measurement of indicators. Figure 2 presents the main components of the European Data Market Monitoring Tool.

Figure 2 Preliminary design of the European Data Market (EDM) Monitoring Tool
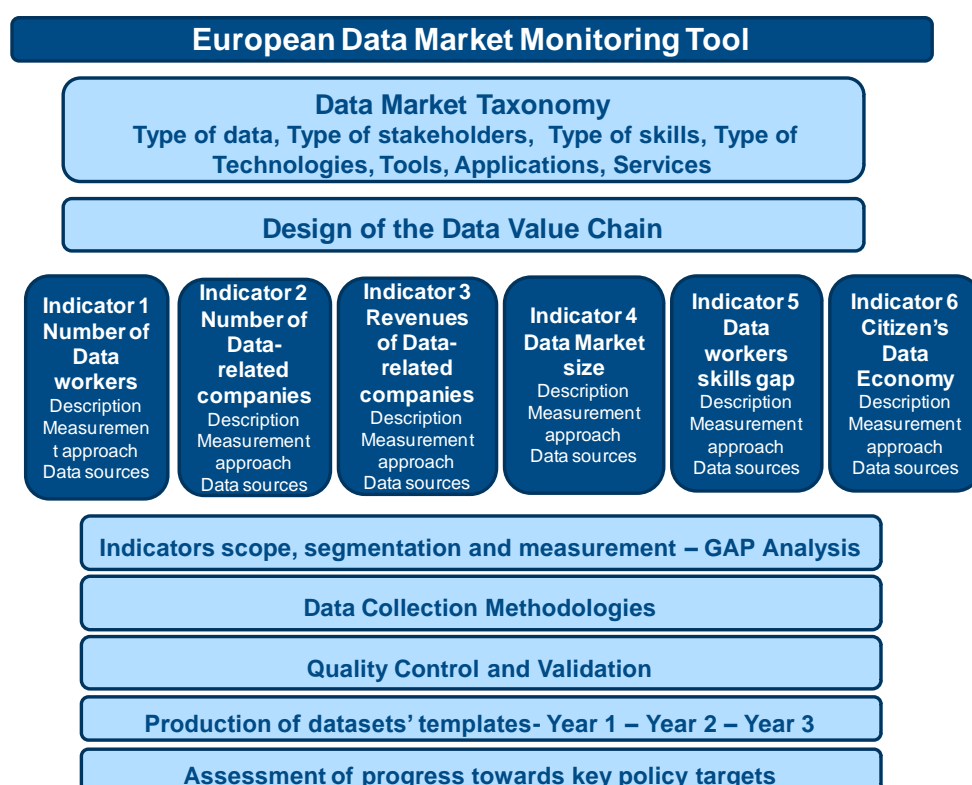
Source: IDC 2014

The monitoring tool has a modular structure, sufficiently flexible to adjust to the market evolution. Its main components are:

- Data Market Taxonomy, introduced above.

SMART 2013/0063 European Data Market

- **Data Value Chain**, introduced above.
- **Framework conditions.** The framework conditions identify the main factors which will enable or prevent the development of the European data market and economy. As indicated in our Data Value Chain design (Figure 1), we have divided the framework conditions into two main groups:
    o Policy/regulatory
    o Market development/non-regulatory

The framework conditions have been identified and classified on the basis of desk research in parallel with the field research activities. The study team has:

    o Reviewed the main existing studies on the data market development conditions
    o Identified the key drivers and barriers to market development, with specific attention to potential bottlenecks due to policy and regulation (either because of legal constraints, or because of lack of action by regulation)
    o Assessed their relative relevance and potential impact on market development based on clear and transparent criteria
    o Identified the potential countermeasures which may reduce barriers, accelerate drivers, and avoid the risk of market underdevelopment
- **Indicators.** The study team has designed and selected a list of indicators. Each indicator is presented through a standardised template, including description, methodological approach, measurement issues, and potential availability of data.
- **Indicator scope, segmentation, and measurement.** A synthetic view of the indicators' scope, segmentation feasibility (by country, sector, and company size), and common issues for measurement and data sources.
- **Data collection methodologies.** For each indicator, the study team has assessed:
    o The availability, quality, and reliability of existing public sources such as Eurostat
    o The availability, quality, and reliability of IDC data
    o The possible combination of the two
    o The study team has further carried out ad hoc field research in a selected group of EU countries/sectors/markets to bridge existing gaps in the available sources
- **Quality control and validation process.** Quality control has been performed on all the steps of the monitoring tool design, development, and implementation by the external experts and by the EC. This will be based on full transparency of the sources and the development process of the monitoring tool.
- **Production of datasets (year 1, 2, 3).** The First Interim Report included the datasets for Year 1; the Second Interim Reports focuses on datasets for Year 2; finally, the Final Study Report included the second datasets for Year 2.
- **Assessment of progress toward key policy targets.** Based on the list of selected indicators corresponding to the main quantitative policy targets indicated by the EC, the study team will assess progress toward policy targets.

3. FORECASTING THE INDICATORS TO 2020

3.1. Three scenarios for the evolution of the data market

3.1.1 Overview

The objective of this chapter is to present three potential scenarios of evolution of the European data market and economy to 2020, based on alternative development paths driven by different macroeconomic and framework conditions. The scenarios assumptions were developed originally in the summer of 2015. They have been revised and updated in February-March 2016 and in October-November of the same year, drawing on the updated data, on a review of the most recent technical and socio-economic developments, and on IDC's worldwide Market Forecast Assumptions, quarterly updated (presented in Annex). As explained further in this chapter, the effects of the Brexit referendum of June 2016 were also taken into consideration in the latest update of our data. The quantitative forecast indicators on GDP and ICT spending have been updated but their correlations driving the quantitative forecast models have not altered radically. The results are presented in the following paragraphs.

The three scenarios provide the storylines, the contextual framework and the main assumptions which have been used to model and forecast the EDM indicators, with a specific focus on the role of policies. Therefore, the scenarios and the forecast models were developed in parallel, testing their relative coherence and fine-tuning their results.

These are medium-term scenarios, with a time horizon of only four years from now, meaning that the range of potential variations of structural factors such as GDP growth and demographics is relatively narrow. Based on these elements, we have established the following scenarios:

- A Baseline scenario was developed first, with the main assumptions based on the continuation of current growth trends and evolution of current framework conditions;

Then we explored the potential alternative development trajectories, resulting in two additional scenarios:

- A High Growth Scenario, where the data market enters a faster growth trajectory, thanks to more favourable framework conditions;
- A Challenge Scenario, where the data market grows more slowly than in the Baseline scenario, because of less favourable framework conditions and a less positive macroeconomic context.

Scenarios are not predictions but potential development paths: their value added lies especially in thinking through the potential consequences of different market trajectories and therefore providing a guide to action. These market scenarios help us to identify the combination of factors and policies best suited to accelerate the development of the data-driven economy in Europe and the size of potential economic gains; and conversely to identify the main challenges to this potential growth.

3.1.2 Scenario Model and Identification of Key Factors

The scenario model used in this study is based on the definition of alternative assumptions about four main groups of key factors driving the data market along different development paths. The identification of the key factors of market development was based on the desk and field research carried out in this study and on the review of a long list of forecast assumptions, leveraging IDC's periodically updated Market Forecast Assumptions (presented in the Methodology Annex). The selection of the most relevant factors 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 four 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, 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 3). The scenarios are characterized 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 3 Structure of the Scenarios Model
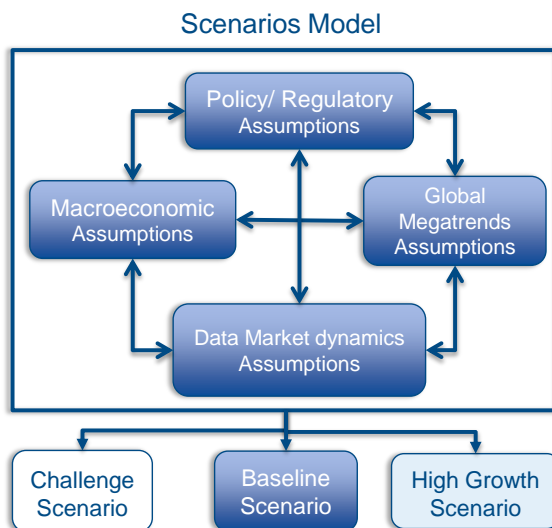
Source: European Data Market Monitoring Tool, IDC 2015

The scenario model and the forecast indicators models are correlated.

Table 2 below summarizes the rationale of their selection and how their assumptions were used as inputs to the indicators' forecast models.

SMART 2013/0063 European Data Market

Table 2 Identification of main factors driving the scenarios

Key Factors	Rationale	Inputs to the forecast models
Macroeconomic factors	Strong influence of the macroeconomic context on data market growth	Alternative forecasts of: EU GDP growth 2014-2015-2016 and 2016-2020 ICT spending growth 2014-2015-2016 and 2016-2020 Alternative economic growth conditions
Policy/Regulatory conditions	Strong influence of the policy/regulatory framework on the model of development of the data market	Alternative policy and regulatory conditions by scenario
Data Market dynamics factors	Strong influence of alternative supply-demand dynamics on the market development paths	Alternative supply and take-up models by scenario
Global megatrends	Strong influence of global digital innovation trends on the EU data market growth	Alternative assumptions on the development of IoT, Cloud Computing, Mobile technologies based the latest on IDC's 2020 forecasts.

The scenarios provide the main framework for the forecast of the EDM indicators. As shown in the figure below, IDC developed seven forecast models: each model produced the specific indicators forecasts under the three main scenarios, followed by in depth cross-check and quality check.

The forecasts models are also correlated and were developed with the following process, with the following dependencies:

- The data market forecast model is the cornerstone of the process: it was developed first, building on IDC's forecasts and on the macroeconomic variables as described below. Its results and growth rates feed into the other models, according to the specific assumption and calculation methods explained for each indicator.
- The data market and data companies'/data users' forecasts influence the data workers model.
- The data companies' forecasts feed into the data revenues model.
- The data workers model feeds into the data workers' skills gap model.
- The data economy model feeds from all the other forecasts, but especially the data market and the data users' forecasts.

Figure 4 Overview of Forecast Indicators Models
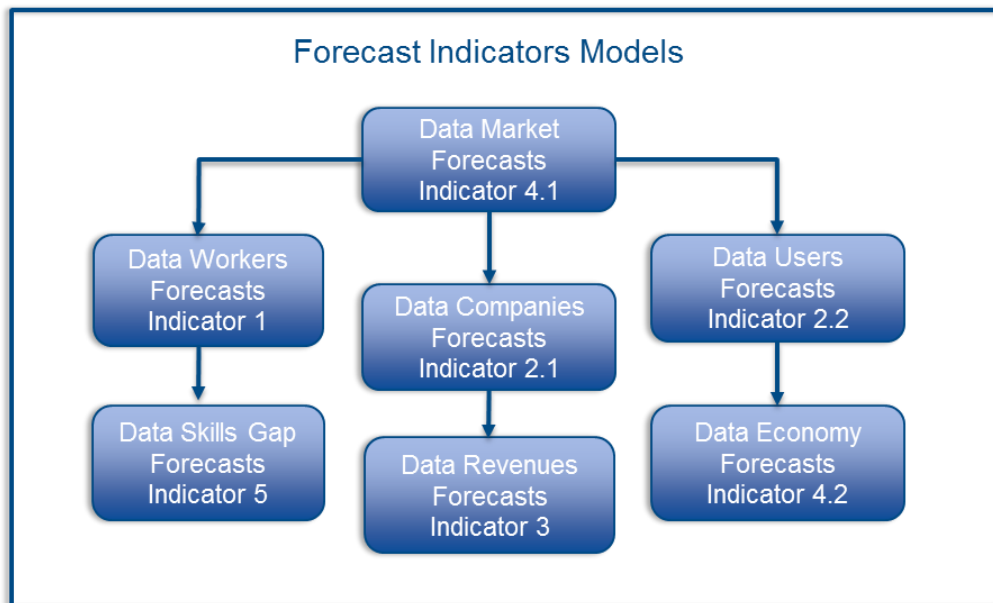
Source: European Data Market Monitoring Tool, IDC 2016

3.1.3 The Brexit Impact on the 2020 Scenarios

In this report we have incorporated the analysis that IDC has recently developed to understand the future developments of the ICT market in Europe taking into account the result of the UK referendum of 23rd June 2016, and the potential impact on the EU data-driven economy as a whole[1]. We are only in the very early stages of what will be a protracted period of change, since the negotiations between the UK and the EU have not even started formally. The following considerations about the impacts of Brexit on EDM's three scenarios must be considered as assumptions and will need to be revised and fleshed out as events evolve and the actual terms of the separation become clearer.

The starting point was the document produced by IDC in July 2016 about the alternative trajectories of IT spending by 2020 following the U.K vote to leave the European Union:

- **Path 1: "Challenging Transition**" — 70% probability. This scenario would see a decline in U.K. GDP at first, but a new relationship set up in some form of bilaterally negotiated agreement in the medium term. Overall we would expect the IT forecast to be revised downwards by more than 2% through to 2020 on a compound annual growth rate (CAGR) basis. Western European IT spend would be expected to remain fairly stable.
- **Path 2: "Disruptive Transition"** — 20% probability. This is most pessimistic scenario and assumes contagion in terms of multiple referenda and immense pressure on the EU model, creating further economic uncertainty. IT spend in this scenario would be expected to decline significantly in the short term and would struggle to rebound in the forecast period in the U.K. and Western Europe. Overall we

---

[1] The Brexit Impact on IT Spend in the U.K. and Western Europe: A Scenario Analysis Insight (Doc #EMEA41570216 / Jul 4, 2016)

would expect the forecast to be revised downwards by close to 5 % through to 2020 on a CAGR basis.

- **Path 3: "Swift Transition"** — 10% probability. This assumes strong leadership steps into the existing vacuum and an orderly Brexit process occurs that avoids short-term turmoil and drives economic growth for the U.K. in the medium term. IT spend is affected mildly in the U.K. in 2016, but rebounds quickly in 2017 and beyond. Europe IT spend unaffected.

In the wake of Brexit, based on recent feedback from a number of large enterprise leadership teams, IDC expects a "wait and see" approach as the political and economic lines are redrawn. IT spending will likely shift, but the strategic transition towards the digital enterprise will remain, and in fact is likely to accelerate with a greater focus on cost optimization and IT value to the organization's bottom line.

Building on these analyses we developed the following considerations which feed into the scenarios:

- The main effects of Brexit on IT markets and the data market are likely to be on the UK itself, while effects on the other EU member states are likely to be more diluted;
- The main economic observers (Consensus Economics[2], IMF[3], OECD[4]) have reduced their future GDP growth forecasts for the UK after Brexit, mainly because of uncertainty undermining investments and demand, even if short term growth has kept up better than expected immediately after the vote;
- The overall macroeconomic scenario for the EU is also expected to be slightly less positive because of losing momentum from one of its largest economies, which had positive growth trends, and because of the long negotiation expected. On the other hand, there could be a positive stimulus coming from the US, if the incoming President Donald Trump actually increased public investments and provided a boost to the North American economy.
- Concerning ICT spending in the UK, IDC expects a slight drop in 2017 and 2018, but demand would recover in 2019 and 2020, and the U.K. ICT market would return to its pre-Brexit levels during 2020;
- The main potentially negative impacts should hit the UK financial services industry (including needs of relocation as well as re-organization) and manufacturing (potential disruption of globalized supply chains). Conversely, these may be gains for other Member States where these activities may be relocated.
- The EU will remain the largest market in the world even without the UK, so global foreign investment flows are likely to remain relevant.
- Finally, the data market growth is driven by long-term trends such as technical and organizational innovation which are independent from Brexit: therefore, we do not expect this event to depress substantially the long-term demand trends, either in the UK or elsewhere;
- On the other hand, the macroeconomic conditions have worsened and this is likely to dampen the potential rate of the data market growth in the next few years, more in the UK than in the other EU MS.

---

[2] http://www.consensuseconomics.com/

[3] World Economic Outlook, October 2016 http://www.imf.org/external/pubs/ft/weo/2016/02/index.htm

[4] OECD Economic Outlook http://www.oecd.org/eco/outlook/economic-forecast-summary-united-kingdom-oecd-economic-outlook-november-2016.pdf

In summary, the new macroeconomic climate following Brexit results in the following variations with respect to the scenarios that we originally envisaged:

- The main storylines of the three scenarios do not change since they focus on demand-supply interactions and innovation trends whose main drivers are not altered by the recent events;
- The GDP forecasts have been reduced compared to the previous scenarios release for the Baseline and High Growth scenarios; the Challenge scenario is the least changed because to some extent it already included negative assumptions. Variations are stronger for the UK than for the other EU27.
- Overall ICT spending is influenced by the macroeconomic uncertainty but also by the combination of opposing trends, the decline of spending in traditional IT and the increase of spending in new technologies which are however more productive and efficient (Figure 5). This leads us to adjust downwards our 2020 forecasts for ICT spending value in 2020 in all scenarios.
- The 2020 data market forecasts have been revised downwards in the baseline and high growth scenario but only by approximately 5% (compared to a minus 20% for ICT spending) since this is an emerging market with strong momentum.
- The 2020 data market value in the challenge scenario was decreased of 2% because it already included pessimistic assumptions about overall growth.
- The number of data companies is substantially unchanged in all the 2020 scenarios: this because none of the new developments affect negatively the number of new companies entering the data market as suppliers. In fact, less competition from the UK which included the highest number of data companies might even be positive for some other MS industry.
- The 2020 forecast number of data users has not changed for the challenge and baseline scenarios, but has been reduced in the high growth scenario due to a smaller number of SMEs buying data market products and services. The assumption is that the lower growth of the data market in 2020 will mostly come from lower SMEs take-up (since the demand by large enterprises is expected to be more resilient and driven by the need to compete with their peers).

3.1.4 Macroeconomic Factors: Background and Relevance

Background and Relevance for the European Data Market

The macroeconomic context will have a strong direct impact on the pace of development of the data market, influencing the availability of risk capital, the amount of investments, and the willingness to spend on new products and services. We expect the data market in the next five years to be very sensitive to the pace of economic growth, because of its early phase of development and the need of investments and investors' confidence. Another important factor is the pace of growth of ICT investments, which is correlated to GDP; this factor is, to some extent, counter-cyclical, as ICT is used by enterprises to improve their efficiency and cut costs, even during a recession. As a result, the amount of ICT spending tends to diminish at a slower pace during a crisis than other types of investments, and is likely to bounce back faster when the recovery occurs. The diffusion of innovative data technologies is positively correlated with overall ICT investments, which include complementary technologies both traditional (servers, network infrastructures) and innovative (cloud computing, mobile and social technologies). However, the pace of growth of the data market is currently faster than the growth of the overall ICT market, which includes a large share of spending for traditional, mature technologies. According to IDC's forecasts, investments in data technologies will continue to outpace traditional ICT investments in the next five years, as enterprises develop their

36

data supply chains and their data-related products and services. As the data market grows in size and data-driven innovation deploys its benefits, its contribution to economic growth will become more visible generating a virtuous cycle of development.

In addition to the qualitative assumptions, we have developed alternative estimates of GDP growth and ICT investments under the three main scenarios, for the EU and each of the EU28 Member States. We used these variables to cross-check and fine-tune the forecast estimates of the data market growth. In other words, this is not a deterministic model where the growth rate of GDP automatically generates a specific growth rate of the data market. Rather, the model took into consideration the coherence of correlation between these variables growth rates.

Figure 5 Main trends of Worldwide ICT spending to 2021
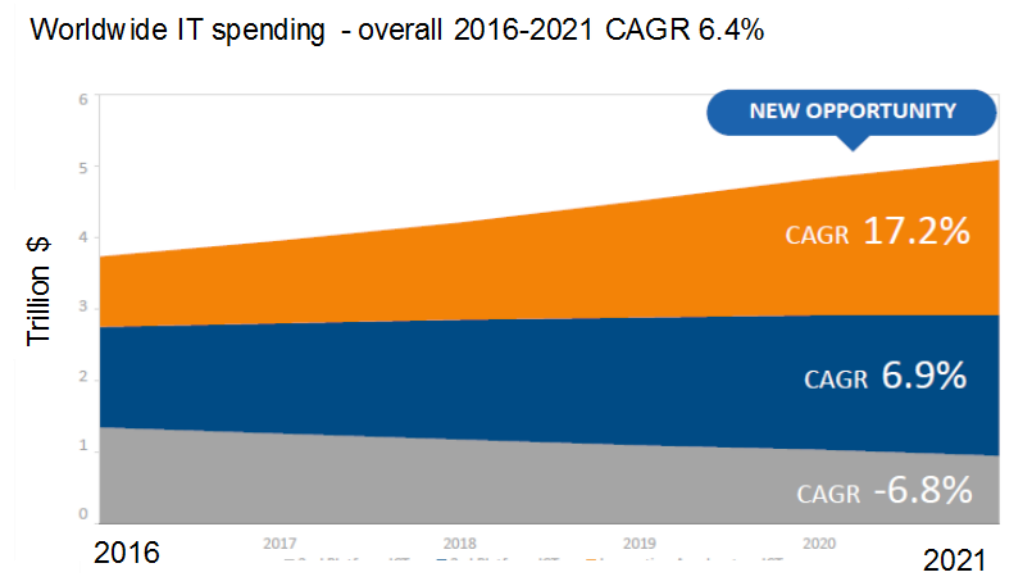
Table 3Legend: Grey area: traditional IT; Blue area: spending in 2nd platform technologies, Cloud, Big Data, Social Media, Mobile; Orange area: spending in innovation accelerators, IoT, Cognitive Systems, Robotics and others

Source: IDC 2016

Tables 3 and 4 present our latest forecast value and growth rates of EU GDP and ICT spending. They were developed for the three scenarios as follows:

- The EU GDP value was estimated on the basis of EIU and EC forecasts (sourced in September 2016).
- The forecast EU GDP growth rate for the period 2016-2020 ranges between 1% for the Challenge scenario, 3.6% for the Baseline and 5.2% for the High Growth scenario. The divergence between the three alternative growth paths is similar to the one estimated in the previous releases of the scenarios, even though the absolute level of the forecast GDP in 2020 in the Baseline and High Growth scenarios is lower, for the reasons explained above.

SMART 2013/0063 European Data Market

- The Baseline scenario value of ICT spending was sourced from IDC's Black Book[5] - 3d Quarter 2016; the challenge and High Growth Scenarios' values were estimated leveraging IDC's database historical series of ICT to GDP correlations. The ICT spending value has been revised downwards because recent evidence collected by IDC shows an accelerating decline of traditional IT spending, partially compensated by higher spending in new ICTs such as IoT, cognitive systems and robotics (Figure 5). The introduction of automation is improving cost efficiency and productivity and therefore decreasing overall IT budgets. In addition, the estimates were revised compared to the last release to adjust for the euro - dollar exchange rate variations in the last two years.
- The combination of these trends results in a forecast decrease of the share of ICT on GDP in all 3 scenarios, from 4.1% in 2016 to 3.4% in the Challenge scenario, 3.7% in the Baseline, 4% in the High Growth scenario.
- However, it is important remarking that enterprise investments in digital innovation are actually growing, but are driven by other business units: according to IDC's 2016 survey, 55% of the investments in digital transformation are made by line of business managers (operations, marketing, human resources…) while only 45% are made by the IT business unit. This means that there is no scarcity of investments in the data-driven economy.

The main macroeconomic variables, their absolute values and growth rates used in moulding the forecast model are summarized below.

Table 3 Macroeconomic Variables used for the Forecast Model – Absolute Values

Macro economic variables	2013 (2013 €-$ exchange rate)	2014 (2014 €-$ exchange rate)	2015 (2014 €-$ exchange rate)	2016 (2016 €-$ exchange rate)	2020 Challenge scenario	2020 Baseline scenario	2020 High Growth scenario
EU ICT spending (€M)	582,285	585,358	619,844	625,347	538,132	636,332	738,878
EU GDP (€M)	13,518,092	13,957,764	14,692,927	15,076,257	15,687,970	17,348,822	18,496,671
ICT/GDP %	4.31%	4.19%	4.22%	4.15%	3.43%	3.67%	3.99%

Legend: GDP growth, EUR Million, at constant exchange rates (2016) and current prices (includes growth of production and prices)

Source: European Data Market Monitoring Tool, IDC October 2016

---

5 IDC Black Books' series are the industry-standard study on the state of ICT spending in every region around the world. IDC's Black Books present a quarterly analysis of the size and growth of the worldwide ICT industry in 54 countries. https://www.idc.com/promo/customerinsights/blackbooks

Table 4 Macroeconomic Variables used for the Forecast Model – Growth Rates

Macroeconomic variables	Growth 2014/2013	Growth 2015/2014	Growth 2016/2015	Challenge Scenario CAGR 2020/2016	Baseline Scenario CAGR 2020/2016	High Growth Scenario CAGR 2020/2016
EU ICT spending	0.53%	5.89%	0.89%	-3.7%	0.4%	4.3%
EU GDP	3.25%	5.32%	2.61%	1.0%	3.6%	5.2%

Source: European Data Market Monitoring Tool, IDC October 2016

3.1.5 Policy/Regulatory Conditions: Background and Relevance

The potential role of policies in meeting the main challenges of development of the emerging data market and data economy has been discussed in depth over the past few years, within and outside the EC. The OECD for example has strongly influenced the debate by identifying supply-side, demand-side and societal challenges requiring policy action (summarised in the following Table 5). The OECD analysis is strongly focused on the potential structural changes of the economy and society brought by data innovation in the medium-long term, and the relative potential threats. The EC and national governments' policy elaboration, instead, has been more focused on the pragmatic need to support and accelerate the adoption of data-related innovation in Europe, recognizing the existence of potential market and regulatory barriers to be removed. This is relevant for our scenarios where policy action is needed to smooth the way towards a well-balanced data ecosystem in Europe. This paragraph focuses on the identification of the policy initiatives with the most direct impact on our forecast scenarios, and a high degree of uncertainty, leading to different policy assumptions for each scenario. To do so, we have focused on the EC-driven policies, as they cover all the main policy issues relevant for the data market and inspire national governments actions.

Table 5 Mapping the main development challenges of the data-driven economy, OECD

Supply- Side Challenges	Demand-Side challenges	Societal Challenges
Need for investments in mobile broadband	Lack of skills and competences in data management and analytics	Potential loss of autonomy and freedom because of mass surveillance and discrimination enabled by data analytics
Limitations to the free flow of data due to regulatory barriers	Need to implement organizational change	Risk of market concentration and dominance in data value chains («winner takes all»)
Data access, ownership and incentives issues for data sharing and re-use	Need for Digital Entrepreneurship to develop data-driven services and products	Risk of greater information inequalities leading to market power imbalances (between organizations, between citizens and governments, between consumers and suppliers)
Access to analytics and cloud computing potentially prevented by lack of interoperability and risks of vendor lock-in		Structural change in labour markets leading to jobs gains but also potential jobs loss, due to the automation of knowledge-intensive and intelligence-based tasks
		New security challenges enabling sharing massive volumes of data

Source: OECD, Data-Driven Innovation for Growth and Well-Being[6], October 2015

---

[6] http://www.oecd.org/sti/ieconomy/data-driven-innovation.htm

In the EC context, different strands of policy elaborated in the last years about data-related challenges (including the Open data and the Data Value Chain policy) were brought together and systematized in the EC Communication "Towards a thriving data-driven economy" (July 2014), defining a coherent Action plan, broadly consistent with the OECD analysis but with different priorities. The Action Plan outlines also the interactions and synergies with other policies and initiatives, such as the Horizon 2020 (H2020) for research and innovation and the Digital Entrepreneurship policy. The Digital Single Market strategy (DSM) launched in May 2015 confirmed the relevance of this issue, targeting the growth of the data-driven economy as one of its key priorities. Some of the actions of the DSM Roadmap are designed to pursue key policy objectives anticipated by the data-driven economy strategy.

More recently, a new set of initiatives were launched by the EC to further sustain the data-driven economy in Europe. In April 2016 the EC published a new Communication on the digitization of the European industry aimed at improving the digitisation process across several economic sectors in Europe[7]. The digitisation of the European Industry would serve as coordination platform for other European, national & regional initiatives (such as Industrie 4.0 in Germany, Smart Industry in the Netherlands or the Nouvelle France Industrielle in France) and is accompanied by other initiatives in the field of standards[8] (to promote widely accepted standards in priority areas such as 5G, Cloud Computing, Internet of Things and Cybersecurity), accelerating the digitisation of public services[9] through a rejuvenated eGovernment Action Plan and reinforcing the uptake and strengthen the benefits to be derived by Cloud technologies[10] in a data-driven economy. At term, all these initiatives will produce results that need to be measured at regular intervals to gauge their progress towards their specific objectives and take corrective actions if necessary.

The data-driven economy strategy, the DSM strategy, as well as the initiative on the digitization of the European industry represent coordinated policy portfolios where the interaction of the main actions is needed to achieve the overarching goals through a systemic impact. They include promoting favourable framework conditions for the data market development through a mix of direct and indirect measures (such as guidelines) and regulation measures. From this point of view, all these policies are relevant: but to consider all their potential interactions in the scenarios would be unfeasible. However, there are different levels of uncertainty about their potential implementation: some are more advanced, others face higher barriers. Also, some policy measures have a long-term focus (such as R&D investments) and are likely to deploy their main impacts after the time period of our scenarios (2020). This is why we used these criteria to select a limited set of policies, whose different outcomes may differentiate the scenarios. The results of the assessment of policy actions by impact and level of uncertainty are presented below. Table 7 presents the key policies selected to elaborate the scenario assumptions. (The assessment of the complete action plan is presented in the Methodology Annex). Table 8 presents the same for the DSM Roadmap actions.

---

[7] Digitising the European Industry – COM(2016) 180 final

[8] ICT Standardisation Priorities for the Digital Single Market – COM(2016) 176 final

[9] EU eGovernment Action Plan 2016-2020, Accelerating the digital transformation of government – COM(2016) 179 final

[10] European Cloud Initiatives – COM(2016) 178 final

Action 1 Community building	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
3. Developing a skills base	High	High. The initiatives are likely to be implemented but whether they will succeed in training a relevant number of skilled data professionals making a difference for the market is uncertain.
Action 2 Developing framework conditions	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
1. Fostering Open Data policies	High	High: uncertain success of the policies in terms of adoption and take-up of guidelines and services
3. Supporting new open standards	High (see also DSM roadmap)	High uncertainty about potential success (difficult process requiring private initiative)
Regulatory issues	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
1. Personal data protection and consumer protection	High	High: strong differentiator between the scenarios depending on the implementation process
3. Security	High	High: complex process, just started, uncertainty about timing and take-up by private sector
4. Ownership/transfer of data	Potentially high, but emerging issue driven by new business models	High uncertainty about timing and impacts (see also DSM roadmap)

Source: IDC elaboration on EC Communication "Towards a thriving data-driven economy"

Table 7 Additional EU Initiatives and level of impact/uncertainty at 2020

Additional Initiative	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
Digitizing the European Industry	High	High. Difficulties in coordinating existing national- and regional-level initiatives and while obtaining buy-in and commitment from the private sector.
ICT Standardization Priorities for the DSM	High	High: uncertain success of the policies in terms of adoption and take-up of guidelines by all MS
Digitalizing Public Services	High	High: existing disparities in digitization capabilities and ICT skills among Public Administrations across the EU
European Cloud Initiative	High	High: Difficulties in obtaining wider access and building trusts among economic operators in Europe

Source: European Commission 2016 and IDC elaboration.

Table 8 Main DSM Strategy Policy Actions by potential impact and uncertainty of outcome

Annex: Roadmap for completing the Digital Single Market Actions	Aspects Relevant for EDM	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
Key Action 1) Better access for consumers and businesses to digital goods and services across Europe	Cross-border online purchases of digital content, cross-border data mining	Medium	Medium –process started, Inception Impact Assessment published in October 2016
Key Action 2) Creating the right conditions for digital networks and services to flourish	Review the e-Privacy Directive (2016)	High, but main impacts after 2020 – no differentiator	High – process started, Inception Impact Assessment published in October 2016
Key Action 3) Maximising the growth potential of the Digital Economy	Initiatives on data ownership, free flow of data (e.g. between cloud providers) and on a	High	High uncertainty about timing and impacts - process started, Inception Impact Assessment

	European Cloud		published in October 2016
	Adoption of a Priority ICT Standards Plan and extending the European Interoperability Framework for public services	High	High uncertainty about actual impact on market - process started, Inception Impact Assessment published in October 2016

Source: IDC elaboration on EC Communication "A Digital Single Market Strategy"

In summary, the policy actions presented in the table below have both a high impact and high level of uncertainty of potential outcomes by 2020 and have been used to define alternative policy assumptions for the scenarios. Actions with a similar scope between the data-driven economy strategy and the DSM strategy are presented side by side, as they must be considered together.

Table 9 Summary of data policy actions used to develop scenario assumptions

Community building	Summary of Actions
3. Developing a skills base	Design a European network of centres of competence to increase the number of skilled data professionals Promote the recognition of new e-infrastructure professions and skills, in line with the 'Grand Coalition on Digital Skills and Jobs' initiative. Make human capital ready for the digital transformation by improving dialogue with social partners, revamp the Grand Coalition for digital jobs as well as the new skills agenda for Europe and promote training within Digital Innovation Hubs.
Developing framework conditions	Summary of Actions
1. Fostering Open Data policies	The EC is preparing guidelines on recommended standard licences, datasets and charging for the re-use of documents; releasing EC and other EU bodies documents as open data through the EU Open Data Portal; creating a pan-European open data digital service infrastructure under the Connecting Europe Facility programme as a one-stop shop; setting up measures and initiatives to promote the open access to research and scientific data sets and sector-specific data (transport, environment).
3. Supporting new open standards	Support the mapping of standards for several big data areas (e.g. smart grid, health, transport, environment, retail, manufacturing, financial services). Identify other sectors with sufficient homogeneity to encourage the further development of standards.
DSM Action	Adoption of a Priority ICT Standards Plan and extending the European Interoperability Framework for public services
Regulatory issues	Summary of Actions
1. Personal data protection and consumer protection	Launch of the GRDP; provision of guidance on its implementation; support of R&I on privacy by design technologies and data sharing tools; launch of a consultation on cloud-based personal data spaces. Support information about and implementation of consumer and marketing law principles in the big data technologies field.
3. Security	Analysis of emerging security risks; proposition of big data related risk management and mitigation measures, including guidelines; support of R&I on solutions reducing data breaches risks and unlawful exploitation of databases.
4. Ownership/ transfer of data	Analysis of the barriers to the cross-border flow of data; of emerging issues of data ownership and liability of data provision, especially for data collected through IoT.
DSM Action	Initiatives on data ownership, free flow of data (e.g. between cloud providers) and on a European Cloud

3.1.6 Data Market Dynamics' Factors: Background and Relevance

The pace of growth of the European data market in the next years will be strongly influenced by Europe's capability to develop a healthy supply-demand ecosystem. This paragraph focuses on the key factors which may lead to alternative supply-demand dynamics correlated with faster or slower data market development paths.

As with most innovative markets, the data market was launched by the push of innovative technology offerings (technology push), but will reach its full potential only when demand is sufficiently mature (demand pull). The analysis of the current data market in Europe, presented in this report, shows a dynamic supply-side (an emerging data industry) and a not-yet fully developed demand-side, where actual users are still a small minority of potential users. Technology-push is still the dominant model of supply-demand interaction.

The growth of demand depends on the penetration of new technologies in the potential users' population, which may be very fast if the innovation is easy to adopt: consider the extremely fast rates of take-up of smartphones. But this study is focused on data technologies, which require process innovation and new skills to be adopted and exploited. The more competitive and innovative enterprises (pioneers) are already adopting them; the more cautious and traditional business users (mainstream adopters) are taking time to enter the data market. The high share of traditional (non-high tech) SMEs in the European economy may contribute to determine a slow demand growth. Traditional SMEs lack awareness of the potential benefits of data-driven innovation and may have difficulty in finding and adopting the appropriate data technologies. As a result, the scenario assumptions about demand dynamics must take into account the factors influencing mainstream adopters and the role of SMEs.

Based on our extensive experience of emerging markets forecasts, we foresee the following potential evolution paths of supply-demand dynamics in the scenarios:

- In the Baseline scenario, we anticipate a healthy growth of the data industry, a continuing improvement of data-driven products and services offerings on the market, and a corresponding gradual development of demand, especially by the most advanced, competitive and innovative enterprises, large and small. Nonetheless, advanced enterprises are a minority of the population of potential users, and in this scenario we foresee only a slow growth of take-up by mainstream, traditional enterprises. In this scenario the supply-demand interaction is still strongly dominated by the supply push.
- However, a High Growth scenario is possible, if take-up accelerates and the adoption of data technologies spreads beyond advanced users to a wide share of mainstream adopters within the next five years. In this scenario the supply-demand interaction moves towards demand-pull and the supply-demand ecosystem is fully developed, with positive feed-back loops between demand and supply.
- In the Challenge scenario, the supply-demand dynamics is similar to the Baseline (with a strong role of technology push) but the growth of demand is weaker, and limited to a smaller population of potential users than in the Baseline.

The main uncertainties about the supply-demand development paths concern the demand and not the supply, which is already on a positive growth path. We have therefore selected the following key factors as the most important differentiators of the supply-demand dynamics in the scenarios.

Table 10 Data Market Dynamics Factors driving Scenarios Assumptions

Key Factors	Rationale and Impact	Level of uncertainty
SMEs willingness and capability to adopt data technologies	EU SMEs represent the wide majority of potential users. A change in their willingness to adopt data technologies would have a substantial impact on overall demand	High

	growth in Europe.	
Awareness of benefits of data-driven innovation	There is a widespread lack of awareness of the potential benefits of data-driven innovation, particularly among SMEs. Greater awareness of successful business cases would give a boost to take-up growth.	High
Diffusion of standards and open platforms for data-sharing	Standardization and interoperability are still immature in the data market. Their diffusion would increase the trust and confidence of potential users. The availability of standardized offerings is a key factor to enable adoption by mainstream users.	High
Availability of appropriate skills	This is a key factor both for the development of the EU data industry and for the capability of the user industries to take-up data technologies.	High
Availability of seed and venture capital funding for start-ups and innovative SMEs in the data market	The overall amount of seed and start-up funding in the EU for ICT-related initiatives was only around €700 million in 2013. If more seed and start-up funding is made available in the coming years this could be a big boost for the EU data industry.	High

3.1.7 Global Megatrends: Background and Relevance

Digital innovation is driven by global trends affecting the world as a whole. The combination of Big Data, Cloud Computing, Mobile technologies and Social media is the most powerful driver of change of the economy and the best opportunity for Europe to move back to a growth path. According to IDC, these four technologies together already account for about 29% of worldwide IT spending, but almost 90% of the spending growth. In addition, innovation will be accelerated in the next years by new developments such as Cognitive Systems, Robotics, 3D Printing and most definitely by the IoT, the Internet of Things, whose networks of sensors will generate huge amounts of data and create "smart environments". The interconnection of these technologies is spreading to all industry sectors, pervading and reshaping business processes and leading to the digital transformation of all enterprises, without exception.

Within this cluster of technologies, Big Data plays a special role as the enabler of most of the innovative services and applications being currently developed. Particularly the combination Big Data, IoT and Cloud Computing is highly effective for digital transformation in the business environment. The diffusion of IoT solutions will generate huge amounts of data for real-time processing and predictive analytics, while cloud computing is the delivery channel enabling the transmission of data and the use of remote data-based services to all enterprises, with pay-as-you-go models. The diffusion of mobile and social technologies in turn generates huge amounts of consumer and business data. IDC tracks digital innovation developments worldwide and in Europe. IDC's research is the source of the key factors presented in the table below which have been used to differentiate the scenarios. In addition, we have recently developed alternative IoT and Cloud Computing market development scenarios to 2020, which we have leveraged to develop the EDM scenarios.

Table 11 Global Megatrends key factors driving scenarios assumptions

Key Factors	Rationale and Impact	Level of uncertainty
Diffusion of the Internet of Things (IoT)	The IoT is an emerging market with a very strong positive correlation and complementarity with Big Data. IDC's IoT 2020 scenarios identify alternative development paths influencing the	High

	data market developments.	
Diffusion of Cloud Computing	Cloud Computing is an emerging market with a very strong positive correlation and complementarity with Big Data. IDC's Cloud 2020 scenarios identify alternative development paths influencing the EDM scenarios.	High
Digital transformation in Europe	Digital transformation requires enterprises to employ digital technologies coupled with organizational, operational, and business model innovation to create new ways of operating and growing businesses. Because of digitization, enterprises in all sectors are moving towards digital business models; Big Data and Analytics are one of the key enabling factors of this transformation.	High
Diffusion of mobile and social technologies	Mobile and social technologies generate massive amounts of data. Their pace of diffusion will influence the data market development.	High

3.2. Description of the Baseline Scenario

The Baseline scenario is defined by a continuation of the 2016 positive moderate growth trend of the European economy, creating favourable conditions for investments in digital innovation in general and data technologies in particular. The increasing diffusion of IoT and Cloud Computing will encourage business demand for Big Data technologies, while the nearly universal penetration of mobile and social technologies by 2020 will herald the emergence of a "hyperconnected" society, where consumers will rely on multiple real-time services for their daily life, often supported by data applications. It is also expected that high-speed broadband infrastructures will be available across Europe and will not become a bottleneck for the data market development.

In this scenario, policy will play an important role to support supply, but have a mixed success in promoting demand, an inherently more difficult objective. Policy initiatives will succeed in supporting the growth of the data industry through R&D investments, the support of digital entrepreneurship, and the successful deployment of the contractual Public Private Partnership on Big Data Value (BDVA PPP). The EU will protect trust in the data economy by successfully deploying the General Data Protection Regulation, achieving greater harmonization across the EU and reducing the administrative burden on businesses. On the other hand, the removal of regulatory barriers preventing the free flow of data cross-borders is unlikely to have effects before 2019-2020. The support of pilot projects and innovation spaces for experimenting with data innovation will help advanced and already interested potential users.

This scenario foresees 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. However, advanced enterprises are a minority of the population of potential users, and in this scenario we foresee only a slow growth of take-up by mainstream, traditional enterprises. Therefore, in this scenario the supply-demand interaction is still strongly dominated by the supply push.

3.2.1 Macroeconomic Assumptions

GDP and ICT Spending Growth

The Baseline scenario foresees a continuation of the moderately positive economic trends experienced by the European economy in 2016, with an average cumulative growth rate

45

of EU GDP by 2020 of 3.6% (see Table 4), taking into account the forecast impacts of Brexit.

These economic trends are supported also by the low level of oil prices and by the monetary policy of the ECB for the euro-area.

ICT spending, however, is expected to grow at a much slower pace and exhibit an average cumulative growth rate of 0.4%, resulting from the combination of two opposite macro trends: stagnation or slow growth of spending for mature technologies, including maintenance or substitution of equipment; fast growth for new technologies spending, including IoT, Cloud, Big data and Mobile. Also, according to IDC some of ICT investments are becoming "embedded" in non-ICT projects (for example, by the marketing or manufacturing departments), so that the total spending in ICT may be slightly under-estimated. Overall, in the Baseline scenario digital innovation investments represent an important driver of economic growth.

Monetary Policy

The European Central Bank is pursuing a monetary policy for growth, through the Quantitative Easing (QE) and APP (Asset Purchase Programme), which are preventing an increase in the real interest rates in the euro-area. The QE is expected to increase the effect of low oil prices on domestic demand, support confidence and investments and therefore support economic growth. The ECB is planning to continue to support expansionary policies as long as the euro-area economy will need it, with positive indirect impacts on the rest of the EU. The continuing risk of deflation has moved the ECB to launch a new, more aggressive programme of Quantitative Easing (QE) which reinforces these trends.

Inflation

The ECB monetary policy has not been able so far to stabilize inflation expectations; if, as expected, the monetary policy will continue and positive effects will last, inflation is expected to pick up gradually and the risk of deflation, which may reduce growth, will be averted.

Job creation/ Unemployment

In this scenario, we expect the structural reforms launched or planned by many Member States to succeed, reducing the rigidity of the labour market, while at the same time the recovery of demand will push enterprises to start hiring again. This will gradually support job creation and a gradual but significant decrease of unemployment during the next five years. However, the growth of employment is expected to be weak in the next years.

3.2.2 Policy/Regulatory Assumptions

In the Baseline scenario we expect a positive trend of growth for the R&I investments in Big Data technologies, driven by national initiatives (already launched, for example, by France, Germany, Ireland and the UK) as well as by the EC with private industries through the BDVA PPP. The PPP was officially launched in October 2014 and is expected to activate investments of EUR 2.5 Billion in the period 2016-2020, of which EUR 500 Million contributed by the EC. These investments and several other initiatives in the EC H2020 Programme will contribute to bridge the gap between research and market and support the growth of the data industry in Europe.

Several Member States, as well as the EC are establishing incubators for data start-ups and SMEs. The EC in particular is planning to launch an Open Data Incubator supported by H2020 to help SMEs set up supply chains based on data. These initiatives will help to increase the number of EU start-ups and innovative SMEs active in the data industry.

The availability of good quality, reliable and interoperable data sets is another key objective of the data-driven economy strategy which is an important enabling condition for the development of demand. However, the policy actions related with this objective are expected to have a mixed success in this scenario, as discussed below.

Developing a Skills Base

The data-driven economy strategy anticipates the design of a European network of centres of competence to increase the number of skilled data professionals, which we assume to see implemented in this scenario. We assume that their goal is to train Big Data analysts or data scientists, not generic data workers. The BDVA PPP is also planning to play an important role to step up training in data skills. In this scenario we assume that these efforts will succeed in preparing a limited number of data professionals but will not solve the demand-supply gap. According to our research (see chapter 10), by 2020 the demand of data scientists and Big Data analysts will grow much faster than supply resulting in a potential gap of 769,000 uncovered positions, corresponding to a 9.8% gap, a significant problem for the European data industry.

Initiatives supporting the recognition of new e-infrastructure professions and skills, such as the launch of the new Digital skills and jobs coalition in December 2016, foreseen by the New Skills Agenda of the EC are aimed at the education system and are not likely to have a relevant effect on the data market development by 2020, even if they are extremely important for the development of the data economy beyond that date.

Fostering Open Data Policies

The policy initiatives foreseen include the development of guidelines on recommended standard licences, datasets and charging for the re-use of documents; releasing EC and other EU bodies documents as open data through the EU Open Data Portal; creating a pan-European open data digital service infrastructure under the Connecting Europe Facility programme as a one-stop shop; setting up measures and initiatives to promote the open access to research and scientific data sets and sector-specific data (transport, environment).

The current level of development of Open Data policies and their potential impacts have been analysed in the study "*Creating Value by Open Data*" by Capgemini Consulting on behalf of the EC, presented in November 2015 in occasion of the launch of the European Data Portal[11]. In the same context the Open Data Maturity indicator was presented, showing that the EU28+ have completed just 44% of the journey towards achieving full Open Data Maturity, as defined by the study. On the use and impact of Open Data, the EU28+ countries score lower, respectively 36 and 29. Using the Open Data Maturity model of the EU28+ countries, the market size of Open Data was calculated. For 2016, the direct market size of Open Data is expected to be 55.3 Bn EUR for the EU 28+.

---

[11] https://www.europeandataportal.eu/en/content/creating-value-through-open-data

Between 2016 and 2020, the market size is expected to increase by 36.9%, to a value of 75.7 Bn EUR in 2020[12].

These results confirm the assumption of the Baseline scenario of a growing adoption of Open Data policies in Europe and potentially positive impacts by 2020, but also highlight the remaining barriers and the fragmented EU scenarios with very different levels of maturity by MS.

These initiatives could play an important role in the stimulation of demand and incentivizing the re-use of data sets, but in this scenario we assume that they will have a positive, but limited impact. For example, the deployment of guidelines is useful, but will likely happen slowly with gradual take-up by potential users.

Supporting new Open Standards and Adopting a ICT Standardization Priorities Action Plan

The EC concluded in January 2016 a consultation on the Priority ICT Standards Plan, which was presented in April 2016[13] according to the DSM roadmap. The plan identified key priorities for Digital Transformation and singled out the data market, promising to increase R&D&I investment specifically for data interoperability and standards as of 2016. This will cover areas such as (i) cross-sectorial data integration (e.g. for entity identifiers, data models, multilingual data management, etc); (ii) better interoperability of data and associated metadata. This will also be used to contribute to global standardisation in the field of data. In addition, support for the BDVA to identify missing standards and develop a common reference architecture were included. In this scenario we assume that these actions will be implemented and that the EC will follow up with these priorities. However, the actual adoption of standards is driven by industry. The assumption for this scenario is that the next years will see a gradual diffusion of open standards and the consolidation of existing ones, with an improvement of the availability of interoperable data sets across sectors, but this will happen slowly leaving many gaps. Some of these standards will not be European but global which will require an adaptation effort by the EU data industry.

Guaranteeing Personal Data Protection

In December 2015, political agreement was reached in the trilogue between the EC, the European Parliament and the Council of Ministers on the General Data Protection Regulation (GDPR). On 8 April 2016 the Council adopted the Regulation and on 14 April 2016 the Regulation was also adopted by the European Parliament. The Regulation will become applicable by 2018. This means that the positive effects of the regulation will start to be felt after 2018. The GDPR is expected to reduce the administrative burden for companies by EUR 2.3 Billion annually[14] through the setting up of a one-stop-shop for

---

[12] https://www.capgemini-consulting.com/open-data-value

[13] https://ec.europa.eu/digital-single-market/en/news/communication-ict-standardisation-priorities-digital-single-market

[14] EU Business, 25th January 2012, http://www.eubusiness.com/topics/internet/data-12/

companies operating across countries, i.e. the responsibility of one single authority for all administrative decisions relating to data processing. Reinforced rules on use and consent, on profiling and on the obligations of companies when handling personal data will reinforce trust of citizens resulting in a continuous sharing of personal information as an important input for value-added data services.

Managing Security in a Big Data environment

Data-driven innovation requires an environment where security does not prevent sharing massive volumes of data and a new digital risk-based approach, involving all stakeholders. In this scenario, the EC successfully supports the stakeholders in a process of analysis of the emerging security risks and identification of possible mitigation measures, which helps to reduce fears and increase trust. Voluntary industry guidelines to identify and manage digital risks are prepared and accepted by leading stakeholders, even though they do not become yet widespread in the forecast period. Towards 2020, technical solutions reducing data breaches risks and the unlawful exploitation of databases (supported also by H2020 projects) start to enter the market.

Removing regulatory barriers to a free flow of data

As foreseen by the DSM roadmap, the EC continued work in 2016 on a European 'Free flow of data' initiative that tackles restrictions on the free movement of data for reasons other than the protection of personal data within the EU and unjustified restrictions on the location of data for storage or processing purposes. It will address the emerging issues of ownership, interoperability, usability and access to data in situations such as business-to-business, business to consumer, machine generated and machine-to-machine data. It will encourage access to public data to help drive innovation. In October 2016 the EC published the Inception ex-ante Impact Assessment for information purposes only, providing two main options for action: addressing data location restrictions through a legislative instruments or a soft-law approach[15]. A decision should be made in 2017. In this scenario, we expect this proposed regulation to require two to three years to be defined and approved, and to be deployed only around 2019-2020, with limited impacts within the forecast period.

3.2.3 Data Market Dynamics Assumptions

As anticipated above, in the Baseline scenario the data industry will continue to grow and improve its offerings, while demand will grow gradually and be dominated by the most advanced and innovative enterprises. The supply-demand interaction model will remain strongly dominated by the suppliers. This is based on the following assumptions.

SMEs Willingness and Capability to adopt Data Technologies

SMEs barriers to entry to the data market are relevant, including lack of skills, low investments in ICT innovation, insufficient access to large datasets and enabling infrastructures, less economies of scale. In this scenario we assume that SMEs in high-tech industries will start overcoming these barriers, because of a competitive drive to

---

[15] http://ec.europa.eu/smart-regulation/roadmaps/docs/2016_cnect_001_free_flow_data_en.pdf

innovate, but a large share of EU SMEs in traditional industries will still be unwilling and unable to adopt data technologies by 2020.

SMEs Awareness of Benefits of Data-driven Innovation

European SMEs (again, excluding those active in the ICT sector and high tech start-ups) are generally little aware of data technologies, consider Big Data an opportunity mainly for large enterprises, and are not able to assess the potential benefits for their specific business. In this scenario we assume that SMEs awareness will gradually increase in the forecast period, but will still remain insufficient to trigger large-scale adoption by 2020.

Diffusion of Big Data standards and Open Platforms for Data-sharing

There is a natural evolution towards standardization in all new ICT markets. This is true also for the data market. Our assumption for the Baseline scenario is that the diffusion of open standards, open platforms for data-sharing and the availability of interoperable data-set will increase slowly in the forecast period to 2020, because of the difficulty to establish new models of interactions between actors and to consolidate consensus behind successful solutions. This process will be helped by proactive policies see the previous paragraph).

Availability of Appropriate Skills

In the Baseline scenario, we assume that EU data users will be able to source sufficient skills to manage the adoption and exploitation of data technologies, even if there will be a need for extensive re-training and recruiting data workers from other careers. The organization of multi-skills work teams should help user organizations to deal with organizational change.

We assume that data companies will have greater difficulties to recruit human resources, particularly with high technical skills (data scientists – see also the previous paragraph on developing the skills base). However, the skills gap should not by itself prevent the successful development of the industry in this scenario, given the moderate rate of growth foreseen.

Availability of Seed and Venture Capital

The amount of venture capital in the EU increased strongly in 2014, according to the EVCA (European Venture Capital Association) and the EBAN (European Business Angels Association), even if the amount of seed and start-up funding was only around EUR 700 Million in 2013. This positive trend is expected to continue in the next years, considering the myriad initiatives launching incubators and accelerators. Overall, in this scenario we assume that sufficient seed and venture capital will be available to EU innovators, but with different levels of availability by geography, with the large Member States (UK, Germany, France) at the forefront of this trend.

3.2.4 Global Megatrends Assumptions

Diffusion of the Internet of Things

According to the Baseline scenario recently developed by IDC for DG CONNECT, IoT revenues in the EU28 will increase from EUR 307 Billion in 2013 to more than EUR 1,181 Billion in 2020, covering sales of IoT hardware, software and services. The number of IoT

connections within the EU28 will increase from approximately 1.8 billion in 2013 (the base year) to almost 6 billion in 2020.

Diffusion of Cloud Computing

According to the Baseline scenario recently developed by IDC for DG CONNECT, EU cloud computing revenues are expected to reach EUR 45 Billion by 2020, representing a share of 10.8% of total IT spending (twice as much as the 4.5% of IT spending covered by cloud in 2015).

Digital Transformation in Europe

According to IDC's Digital Transformation Benchmark survey[16] carried out in 2016 on a sample of large enterprises over 500 employees, the majority of EU enterprises are implementing Digital Transformation in an opportunistic way and only 5% can be defined as Digital Leaders, with a slight increase on 2020 (when it was only 3.5%). The level of digital maturity lags behind that of their US peers. Given these data, we assume that Digital transformation by 2020 will engage most large enterprises and a majority of middle-size enterprises, but still a minority of small enterprises.

Diffusion of Mobile and Social Technologies

The almost universal penetration of mobile and social technologies will lead to the "hyperconnected" society by 2020. Mobile devices will multiply the amount of data available for analysing customers and suppliers; online customer engagement will become the normal practice in most industries.

3.3. Description of the High Growth Scenario

In the High Growth scenario, Europe's economic growth in the next years will be higher than the Baseline scenario and will be characterised by a stronger driving role of digital innovation, with higher overall ICT investments as a share of GDP. Solutions combining innovative digital technologies (such as IoT, Cloud and Big Data) will be more widely implemented and more European enterprises will engage in Digital Transformation before 2020. The data market will enter a faster growth trajectory and the adoption of data technologies will spread beyond the minority of pioneers to a wider population of mainstream users. 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. This is a classic virtuous cycle mechanism, which may happen if data technologies take-up starts climbing fast enough to generate momentum. Because of network effects typical of ICTs, rapid diffusion multiplies the benefits for users in their interactions and makes it easier to consolidate standards and interoperability, reducing further the barriers to adoption.

To enable this scenario, we must assume a set of very favourable framework conditions, which will be able to trigger a faster take-up. First of all, the adoption of all digital technologies is mutually reinforcing, so we assume a faster pace of diffusion for IoT, Cloud, Mobile as well as data technologies. Second, we must assume a leap ahead of

---

[16] IDC MaturityScape Benchmark: Digital Transformation in Europe, August 2015, http://www.idc.com/getdoc.jsp?containerId=DTS03X

awareness of potential benefits and willingness to adopt data technologies by mainstream users and specifically by SMEs. Third, but not less relevant, we must assume a removal of existing regulatory barriers within the forecast period. In this scenario, policy initiatives will succeed in supporting supply as detailed above, but will also have better success in promoting demand. Policies enabling the free flow of data cross-borders and the re-use of data sets will create positive effects on demand already from 2017-2018. All the other positive factors described in the Baseline scenario must also be present.

3.2.5 Macroeconomic Assumptions

GDP and ICT Spending Growth

The High Growth scenario foresees an average cumulative growth rate of EU GDP between 2016 and 2020 by 5.2%, higher than in the Baseline scenario (see Table 4). ICT spending instead is expected to increase by an average cumulative growth rate of 4.3% (against 0.4% in the Baseline scenario), representing 4 % of EU GDP in 2020 (against 3. 7% in the Baseline scenario). This increase is driven by a larger share of investments in new technologies such as IoT and Big Data, compared to the Baseline scenario.

Monetary Policy and Inflation

These assumptions are the same as in the Baseline scenario.

Job creation/ Unemployment

In this scenario we assume a slightly faster reduction of overall unemployment, because of the job creation driven by digital innovation. However, there will also be jobs loss in traditional jobs so the net gains will not change radically the balance compared to the Baseline scenario.

3.2.6 Policy/Regulatory Assumptions

In this scenario, policy initiatives will succeed in supporting supply as outlined in the Baseline scenario, but will also have better success in promoting demand as outlined below.

Developing a Skills Base

In the High Growth scenario, the fast growth of demand of data skills compared to the more rigid dynamics of skills supply is expected to generate a higher number of unfilled data workers position than in the Baseline scenario. Assuming that the policies aimed at creating European networks of competence centres training data professionals will be successfully implemented, still they will not be able to satisfy completely the new demand. Therefore, in this scenario we expect the BDVA PPP to step up its training efforts, and both the EC and national governments to launch more proactive policies. The insufficient supply of data scientists will be particularly felt by the supply industry. However, we assume that user industries will find ways to get around the problem, by establishing multi-professional teams or by relying on outsourced skills.

52

As in the Baseline scenario, we anticipate to see the implementation of initiatives supporting the recognition of new e-infrastructure professions and skills, such as the launch of the new Digital skills and jobs coalition in December 2016, foreseen by the New Skills Agenda of the EC, with impacts occurring mainly after 2020.

Fostering Open Data Policies

In this scenario, we expect these policy initiatives to have a positive effect on the stimulation of demand and to provide incentives for the re-use of data sets. The deployment of guidelines for data re-use will prove their usefulness and lead to take-up by users. The use of open data sets published by public portals will increase faster than in the Baseline scenario attracting many SMEs.

Supporting new Open Standards and Adopting a ICT Standardization Priorities Action Plan

In this scenario we foresee the successful implementation of the ICT Standardisation Priorities Action Plan presented in April 2016, followed by the identification of the missing standards and the promotion of countermeasures.

The data industry will drive a fast diffusion of open standards and the consolidation of existing ones, motivated by the need to standardize offerings for a wider population of non-sophisticated users and SMEs. Some of these standards will not be European but global which will require an adaptation effort by the EU data industry.

Guaranteeing Personal Data Protection

In this scenario, we expect the General Data Protection Regulation (GDPR) to enter into force by 2018. It will reduce the administrative burden for companies by EUR 2.3 Billion annually through the setting up of a one-stop-shop for companies operating across countries, i.e. the responsibility of one single authority for all administrative decisions relating to data processing. Additionally, rules on consent of re-use of data for purposes different than the original purpose of collection, and data minimisation will allow Big Data analytics to exploit more data with less restrictions than in the Baseline scenario; implementation of the new regulation will be largely harmonious throughout EU countries, creating the legal certainty needed by companies; certification schemes add to legal certainty by giving companies the necessary certainty and trustmark for exploitation of personal information in new ways, respecting privacy-by-design principles; a revised ePrivacy Directive will produce results earlier than originally expected and create a level playing field for telecoms so that they themselves or their business partners can exploit the full potential of mobile phone data.

Managing Security in a Big Data Environment

In this scenario, as in the Baseline the EC will successfully support the stakeholders in a process of analysis of the emerging security risks and identification of possible mitigation measures, which will help to reduce fears and increase trust. Voluntary industry guidelines to identify and manage digital risks will be prepared and accepted by leading stakeholders and become rapidly widespread within the forecast period. Towards 2020, technical solutions reducing data breaches risks and the unlawful exploitation of databases (supported also by H2020 projects) will start to enter the market.

Removing Regulatory Barriers to a Free Flow of Data

The EC 'Free flow of data' initiative for Europe will be launched in 2017 but in this scenario we expect it to be deployed at an earlier moment in time than in the Baseline scenario. By establishing coherent and harmonized regulations for cross-border use of data, interoperability, usability and access to data issues, the initiative will create a favourable environment for take-up, contributing to the increase of demand.

3.2.7 Data Market Dynamics Assumptions

In this scenario, 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.

SMEs Willingness and Capability to Adopt Data Technologies

In this scenario we assume a positive change in SMEs' willingness and capability to adopt data technologies by 2020, because of competitive pressure to innovate (given the overall faster diffusion of digital innovation) and a higher availability of investments. Also, the quality of offerings for SMEs is assumed to be better in this scenario than in the Baseline. This will result in a higher share of SMEs becoming data users compared to the Baseline scenario.

SMEs Awareness of Benefits of Data-Driven Innovation

In this scenario we assume that the deployment of successful examples of data-driven innovation in the market will prove the business case for SMEs. The dissemination of these examples and this knowledge with the support of public awareness campaigns will generate a leap ahead in SMEs awareness, eventually resulting in higher willingness to adopt data technologies.

Diffusion of Big Data Standards and Open Platforms for Data-sharing

Our assumption for this scenario is that the diffusion of open standards, open platforms for data-sharing and the availability of interoperable data-set will increase rapidly in the forecast period to 2020, driven by a need to satisfy increasing demand. This process will be helped by proactive policies.

Availability of appropriate Skills

In the High Growth scenario there is likely to be a worse problem of sourcing data skills, partially compensated by more intensive re-training and recruiting data workers from other careers, supported by public and private initiatives. The organization of multi-skills work teams should help user organizations to deal with new skills demand and organizational change, as in the Baseline scenario.

In this scenario, too, data companies will have difficulties in recruiting human resources, particularly data scientists, even if the greater attractiveness of data careers will help by attracting professionals from other careers (in the short term) and by encouraging more students to enrol in the courses teaching data technologies and management (in the long term). The data skills gap should not by itself prevent the successful development of the industry in this scenario, but is likely to affect more strongly start-ups and innovative

SMEs, with the risk to slow down their expansion. This may in turn lead to outsourcing, opening a potential market for specialized data services companies.

Availability of Seed and Venture Capital

In this scenario we assume an increase in the availability of seed and venture capital to EU innovators compared to the Baseline scenario, but with different levels of availability by geography, with the large Member States (UK, Germany, and France) at the forefront of this trend.

3.2.8 Global Megatrends Assumptions

Diffusion of the Internet of Things

This scenario foresees a higher diffusion of solutions combining different technologies, such as the IoT, Cloud and Big Data, than the Baseline. IDC's estimates of fast IoT growth in this scenario are of EUR 1,128 Billion of revenues in 2020 (only 5% higher than the Baseline scenario revenues). Since IoT is already expected to grow very rapidly in the Baseline scenario, this small positive gap is understandable.

Diffusion of Cloud Computing

In this scenario, EU cloud computing revenues are expected to reach EUR 60 Billion by 2020, representing a share of 14.4% of total IT spending (against 10.8% in the Baseline scenario).

Digital Transformation in Europe

In this scenario, we assume that a higher share of EU enterprises than in the Baseline scenario will engage in Digital transformation before 2020, including the more innovative SMEs. The faster diffusion of innovative business models will increase competition and emulation among enterprises.

Diffusion of Mobile and Social Technologies

The almost universal penetration of mobile and social technologies will lead to the "hyperconnected" society by 2020. Mobile devices will multiply the amount of data available for analysing customers and suppliers; online customer engagement will become the normal practice in most industries, as in the Baseline scenario.

3.4.   Description of the Challenge Scenario

In the Challenge scenario, the combination of a less positive macroeconomic context than in the Baseline scenario, less favourable framework conditions, and slower diffusion of digital innovation, will combine to push the data market into a low growth development path. This is a fragmented scenario, where the Digital Single Market will fail to

55

materialize before 2020. The supply-demand dynamics will be dominated by the technology push, since the demand pull will be weak. The level of adoption of data technologies by 2020 will be limited to a smaller population of potential users than in the Baseline, as market barriers to entry will remain high. This scenario therefore explores the potential risks and consequences of failing to remove the barriers to the development of the data economy in Europe.

This scenario still anticipates an increase of the diffusion of digital technologies such as IoT and Cloud, but at a slower pace than in the Baseline. The dynamics of mobile and social technologies should not be much different in this scenario, given their strong momentum and their closeness to nearly universal diffusion. As a result, the "hyperconnected" society will become closer in this scenario too, even if less well developed than in the Baseline or High Growth scenarios. It is possible that the diffusion of high-speed broadband infrastructures across Europe will be incomplete, with the risk of a digital infrastructures divide between and within the Member States. This will be another element of weakness for the development of the data market.

In this scenario, both supply-side policies and demand-side policies will tend to have weaker impacts and to be deployed more slowly in time. Policy initiatives will still succeed in supporting the growth of the data industry through R&D investments, the support of digital entrepreneurship, and the successful deployment of the BDVA PPP, but to a lesser extent given the lower propensity to invest by the private sector. Policies addressing enabling conditions, such as the removal of regulatory barriers to the free flow of cross-border data, will be delayed in time and be less effective than in the Baseline scenario. As a consequence, the value of the data market and of the data economy by 2020 will be substantially lower than in the Baseline scenario.

3.2.9 Macroeconomic Assumptions

GDP and ICT Spending Growth

In this scenario, the effects of the economic crisis suffered by Europe in the period 2008-2013 and beyond will continue to be felt, with a negative impact of Brexit and the political uncertainty. The moderate economic recovery started in 2015 will continue, but at a slower pace. The divergence between Member States' economic development paths will increase again, with some countries struggling harder than others with the legacy of the crisis. This context will create uncertainty, potentially reducing trust and confidence and the overall level of investments.

The Challenge scenario foresees an average cumulative growth rate of EU GDP between 2016 and 2020 of 1%, much lower than the Baseline scenario (see Table 4). ICT spending is expected to actually decrease with an average growth rate of minus 3.7% between 2016 and 2020 (against positive growth of 0.4% in the Baseline scenario). The share of ICT spending on EU GDP will decrease to 3.4% versus 4.1% in 2016.

Monetary Policy

There are risks that the ECB Quantitative Easing (QE) will not be sufficient to help Member States with structural imbalances to get back on the path of faster growth. The Greek crisis, still not solved, may have more negative impacts than expected. Oil prices may rebound faster than expected. In this scenario one or more of these risks will materialise.

Inflation

In a more negative economic scenario, lack of confidence by businesses and citizens will slow down the initial (2015-2016) investment and consumption trends. In turn, this will maintain high the risks of deflation in most of the EU Member States, depressing demand and investments in a vicious cycle.

Job creation/ Unemployment

Because of the weak economic recovery, Europe will struggle to increase job creation and bring down unemployment from the pre-crisis levels, even if some employment increase is expected.

3.2.10 Policy/Regulatory Assumptions

In this scenario, policy initiatives will partially succeed in supporting supply, but will have mixed success in promoting demand as outlined below.

Developing a Skills Base

In the Challenge scenario, the policies aimed at creating European networks of competence centres training data professionals will be implemented, but they will receive less support and collaboration from the private industry and will train a smaller number of data professionals than in the Baseline. The insufficient supply of data scientists will still be a problem for the supply industry, but at a lower level than in the Baseline scenario given the less dynamic data market growth.

As in the Baseline scenario, we expect to see the implementation of initiatives supporting the recognition of new e-infrastructure professions and skills such as the launch of the new Digital skills and jobs coalition in December 2016, foreseen by the New Skills Agenda of the EC, with impacts occurring mainly after 2020.

Fostering Open Data Policies

In this scenario, we expect these policy initiatives to be deployed slowly and to achieve a limited effect on the stimulation of demand and the incentives for the re-use of data sets. The deployment of guidelines for data re-use will happen slowly with limited take-up. The use of open data sets published by public portals will improve, but gradually.

Supporting new Open Standards and adopting a ICT Standardization Priorities Action Plan

The ICT Standardization Priorities Action Plan presented in April 2016 will take longer than expected to be implemented, with potential low effects on the data market. The assumption for this scenario is that the next years will see a gradual diffusion of open standards and the consolidation of existing ones in the data market, with an improvement of the availability of interoperable data sets across sectors, but this will happen very slowly leaving many gaps.

Guaranteeing Personal Data Protection

In this scenario, we expect the GDPR to be adopted after 2018 with several delays. The one-stop-shop mechanism will be severely diluted in the implementation process, creating a complex cooperation mechanism that slows down administrative decision-

making and limits the expected reduction of administrative burden. At the same time, rules on consent, data minimisation and re-purposing of use made of personal data are adopted in a way that makes it very difficult for companies to use personal data for new interesting purposes on the basis of Big Data analytics; anonymisation and pseudonymisation are seen with great suspicion by data protection authorities. Interpretations of the new GDPR vary across countries adding to a degree of distrust of citizens. In combination with new data leak scandals, citizens share less and less personal information – even in exchange for 'free services', clearly limiting the uptake of data services building on personal information. The revision of the ePrivacy Directive will not be adopted before 2020, impacting particularly on the use of data held by telecoms.

Managing Security in a Big Data Environment

In this scenario, the EC will find difficulties in promoting the analysis of the emerging security risks and identification of possible mitigation measures with the main stakeholders. Leading stakeholders will disagree on voluntary industry guidelines aimed at defining good practices for managing digital security risks; their conflicts will prevent the development and adoption of guidelines. Technical solutions reducing data breaches risks and the unlawful exploitation of databases (supported also by H2020 projects) may start to enter the market after 2020.

Removing Regulatory Barriers to a Free Flow of Data

The EC 'Free flow of data' initiative for Europe foreseen by the DSP Roadmap is expected for 2017 but in this scenario we expect a later launch and a deployment towards the end of the forecast period because of disagreements between the Member States. Because of the delay in establishing coherent and harmonized regulations for cross-border use of data, interoperability, usability and access to data issues, enterprises will operate in a fragmented environment, suffering from barriers to take-up of data technologies.

3.2.11 Data Market Dynamics' Assumptions

In the Challenge scenario, the data industry will continue to grow, but demand will increase slowly with only the most advanced enterprises adopting data technologies. The supply-demand interaction model will remain strongly dominated by the suppliers. This is based on the following assumptions.

SMEs Willingness and Capability to Adopt Data Technologies

In this scenario, SMEs barriers to entry to the data market will not decrease. We assume that a minority of SMEs in high-tech industries will start overcoming these barriers, because of a competitive drive to innovate, but the majority of EU SMEs in traditional industries will still be unwilling and unable to adopt data technologies by 2020.

SMEs Awareness of Benefits of Data-Driven Innovation

In this scenario we assume that SMEs awareness of potential benefits will gradually, but very slowly increase in the forecast period, and will remain insufficient to trigger large-scale adoption by 2020, even more so than in the Baseline scenario.

Diffusion of Standards and Open Platforms for Data-sharing

Our assumption for the Challenge scenario is that the diffusion of open standards, open platforms for data-sharing and the availability of interoperable data-sets will increase very slowly in the forecast period to 2020, because of the difficulty to establish new models of interactions between actors and to consolidate consensus behind successful solutions. Policy actions will support standardization for the data market, but with indifferent success.

Availability of Appropriate Skills

In the Challenge scenario we expect a lower gap of data skills compared to 2016 and to the other scenarios, because of the lower level of demand. We assume that the few EU data users will be able to source sufficient skills to manage the adoption and exploitation of data technologies, even if there will be a need for extensive re-training and recruiting data workers from other careers. The organization of multi-skills work teams should help user organizations to deal with organizational change.

Availability of Seed and Venture Capital

The availability of seed and venture capital will be lower in this scenario than in the Baseline, because of the less favourable economic context. However, there are still likely to be many initiatives supporting data start-ups (accelerators, incubators), because of the promised growth potential of the data market. The divide between Member States in terms of providing sufficient risk capital to start-ups will deepen again, with only the UK, Germany and France able to provide sufficient public funding for this goal.

3.2.12 Global Megatrends Assumptions

Diffusion of the Internet of Things

Based on IDC's IoT scenarios, in the context of less favourable economic conditions the potential value of the IoT market by 2020 would reach EUR 976 Billion, which is 18% lower than the value in the Baseline scenario.

Diffusion of Cloud Computing

In this scenario, EU cloud computing revenues are expected to reach only EUR 24 Billion by 2020, representing a share of 6.9% of total IT spending (against 10.8% in the Baseline scenario).

Digital Transformation

In this scenario, we assume that Digital Transformation will be taken up only by the more advanced and competitive users, including the more innovative SMEs. Organisational inertia and cultural barriers will slow down the digital transformation of enterprises, particularly those from highly regulated, public services industries: healthcare, energy, government sectors. The overall investments in Digital Transformation will be much lower than in the Baseline scenario.

Diffusion of mobile and social Technologies

SMART 2013/0063 European Data Market

Mobile and social technologies' diffusion will grow, but with lower average spending. The diffusion of the hyperconnected society will be uneven across Europe, with a digital divide between more and less "digitally" mature Member States.

4. MEASURING THE DATA WORKERS

4.1. Definition

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

The identification, definition and measurement of data workers is a recent field of investigation. High-tech systems created a massive quantity of data; therefore, to use and exploit such data, data analytics professionals are necessary.

The data industry is an emerging industry and the use of data is at an early stage. The amount of data which is being produced and used is progressively increasing, driving the development in data-related technologies and the uptake of tools to process this data. The diffusion and adoption of data depend on two factors:

- The availability of technology and tools to collect, process, and analyse data, and make data usable.
- The awareness and capability of users to process and analyse data. It may be that some users, aware of the potential effects of data, analyse it and use it although they "under-use" the technology for data processing and analysis.

In our definition, data workers are not only data technicians but also data users who based their business decisions on their analysis and interpretation of the data. According to our definition, data workers belong to the category of knowledge workers and specifically "codified" knowledge workers (Lundavall and Johnson, 1994); data workers specifically deal with data while knowledge workers deal with information and knowledge.

In the Methodological Annex, we also discuss the recent definitions adopted by the OECD (2015) and the European Data Science Academy, which both adopt a much narrower approach than we have done.

4.2. Measuring data workers

The 2015 data workers measured 6 million in EU28. For the year 2016, we estimate that data workers are increasing with a growth rate 2016-2015 which is nearly 5%, reaching 6.1 million units.

The average number of data workers per user company (i.e. the number of data workers in relation to the number of data users) is stable at around 9 units per company. A growing trend of data workers with a stable number of workers by user company supports the hypothesis that the diffusion of data products and services is gradually spreading.

Table 12 Indicator 1 Data Workers

Indicator 1 — Data Workers 2014-2015-2016								
N.	Region	Name	Description	2014	2015	2016	Growth rate 2015/2014	Growth rate 2016/2015
1.1	EU27	Number of data workers	Total number of data workers in EU (000s)	4,707	4,730	4,941	0.5%	4.5%
1.1	EU28	Number of data workers	Total number of data workers in EU (000s)	5,818	6,005	6,161	3.2%	2.6%
1.2	EU27	Employment share	Share of data workers on total employment in EU (%)	2.8%	2.8%	2.9%	-0.5%	3.5%
1.2	EU28	Employment share	Share of data workers on total employment in EU (%)	3.0%	3.1%	3.1%	2.2%	1.7%
1.3	EU27	Intensity share	Average number of data workers per user company (units)	9.5	9.7	9.8	1.7%	1.6%
1.3	EU28	Intensity share	Average number of data workers per user company (units)	9.1	9.2	9.3	1.9%	1.1%

Source: European Data Market Monitoring Tool, IDC 2016

4.2.1 Updating the Measurement of Data Workers in 2016

As the statistical data to estimate 2016 data workers are not yet available, we remind that the estimates for the year 2016 are a forecast. As explained in detail in the Methodological Annex, this forecast was based on the following assumptions:

- Countries investing in data products experience an increase in TFP (Total Factor Productivity) growth
- Data market growth is a function of data labour growth, stock capital growth, and TFP growth using a production function approach
- The data market total factor productivity is similar to the average total European factor productivity
- The ICT capital stock is equal to the sum of the last three years in ICT investment.

4.3. Indicator 1.1 and 1.2: Number of Data Workers and Employment Share

4.3.1 Data Workers by Member State

Six Member States (UK, Germany, France, Italy, Poland, and Spain: the "Big Six") account for 72% of the total data workers in 2015 as well as in 2016, while the remaining 28% of the data workers of the EU are distributed across the other 22 Member States. The number of data workers depends on the data market trend but also on other factors, which are the ICT stock and the productivity. This explains why some countries, such as Poland for example, count more data workers than Spain although their data market is lower than the data market in Spain. This depends on the fact that in Spain the ICT stock is much higher: in other words, the ICT inputs increases the data workers' productivity so that the workers needed are less than the data workers needed by a country with a lower ICT stock. The share of the data workers on the total employment is stable for the EU28. This share varies significantly by country, going from 6.5% in Luxembourg to 2% in Romania with an average of data workers on total employment which is around 3%. Two countries among the "big 6", Italy and Spain, have a share of data workers which is lagging behind the European average. This relates in part on their ICT spending, and significantly on their industry structure where SMEs are very relevant; as we know, for small businesses the data products and services may be less accessible than they are for large companies. In terms of employment share, the discrepancy between big and small countries tends to lose importance while structural factors, both country and industry specific appear to be more relevant.

Figure 6 Number of data workers by MS, 2016 (000s) and growth rates 2016/2015

Source: European Data Market Monitoring Tool, October 2016

SMART 2013/0063 European Data Market
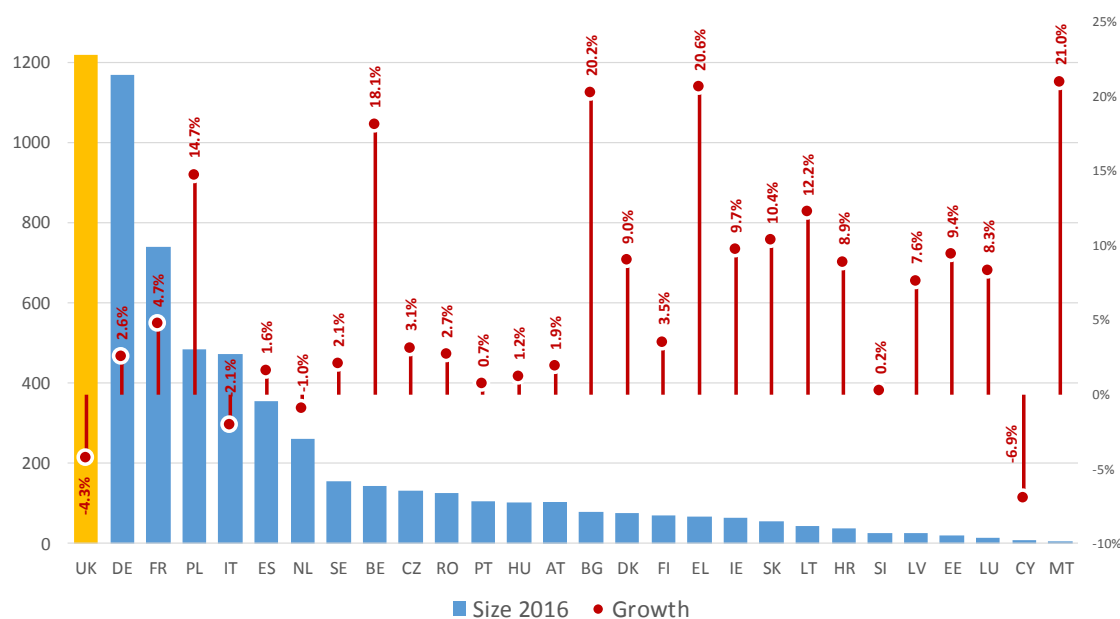
4.3.2 Data Workers by Industry

European data workers are distributed in nearly all industries, but their employment share by industry varies substantially. Four industries — manufacturing, wholesale and retail, professional services, and ICT — represented nearly 62% of data workers in 2016 with no significance differences with the previous years (nearly 3,800 on the overall 6,161). In absolute terms, professional services count for 20% of the population of data workers, followed by wholesale and retail with another 18%, and then manufacturing (12%) and information and communication (11%). However, in terms of the share on total employment, ICT and Finance lead, with professional services in the third place. The industries with the lowest concentration of data workers are Construction, Transport and Healthcare. ICT and Finance represent the industries with the highest level of IT spending and the highest propensity to exploit data. The strong presence of data workers in professional services and retail show the increasing diffusion and relevance of data-driven services in these industries, particularly for marketing and customer services. These industries are also undergoing a deep digital transformation process, using digital technologies to re-invent products and services, forced by competition and disruptive innovator start-ups. Manufacturing and the utilities sectors are also evolving towards digital transformation and the presence of data workers is already quite relevant. The data for the years 2013 to 2016 show similar dynamics of penetration by industry, due to the influence of structural characteristics of the industries on the presence of data workers. Structural change processes take time, and therefore the distribution of data workers among industries is not going to show in the short to medium term important changes.

Figure 7 Number of data workers by Industry 2016 (000s), and Share (%)
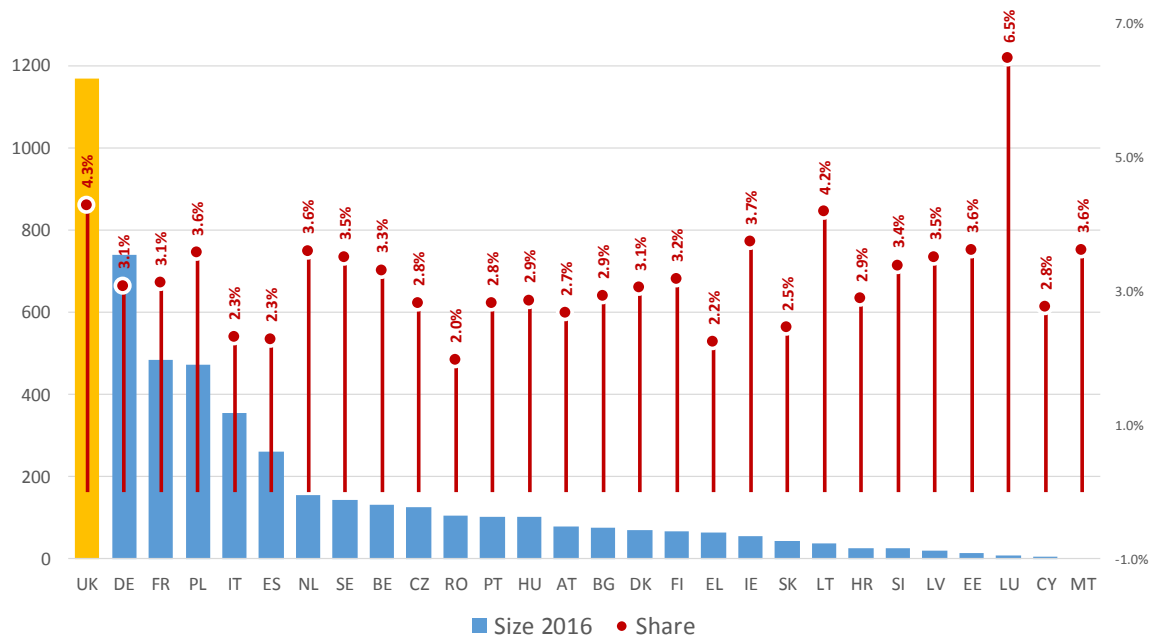
Source: European Data Market Monitoring Tool, October 2016

SMART 2013/0063 European Data Market
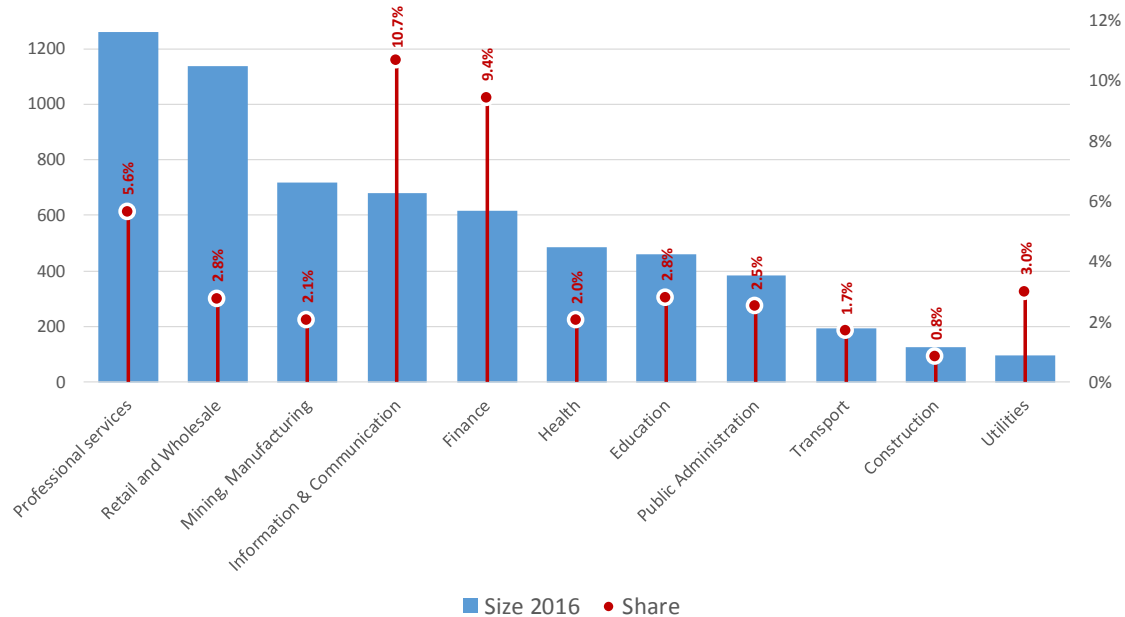
Source: European Data Market Monitoring Tool, October 2016

In 2016, the total number of data workers in the EU28 grew by 2.7% over the last year. On the overall, data workers grew with a trend similar to the total employment trend. The ISCO occupations where the data workers are concentrated did not grow up faster than the general employment and therefore the data workers trend has been similar to the general employment trend.

Although the employment is going to remain nearly flat, we believe the number of the data workers is going to increase faster than the total employment because, based on IDC forecasts, the ICT spending is showing a positive trend which is in part supporting new investments in the data technology.

Data is one of the production factors of the economy; moreover, data is a pervasive and multipurpose production factor. Countries and industries that have invested significantly in data products and services may experience an increase in the overall efficiency of labour and capital or in total factor productivity (TFP) growth.

Industries receive productivity gains from data use over the labour productivity gains received from investments in ICT, for instance because of an improvement in production processes.

4.4. Indicator 1.3: Intensity of Data Workers

4.4.1 Description and Methodology Approach

This indicator measures the average number of data workers calculated on the total number of data user companies. This indicator correlates two other indicators calculated with different models and methodologies: the number of data workers is based on

employment statistics, while the user company indicator builds on the business enterprise statistics for the private sector (excluding government organizations which are not measured as units per Member State by Eurostat). Therefore, they measure two partially different perimeters of the data market.

This means that this indicator should be considered carefully and only as indicative, particularly at MS level. At the EU level the average number of 9.3 data workers per user company in 2016 is stable where compared to the 2015 data (9.2). The growth dynamics of the intensity of data workers mirror the one of the data workers. Based on our estimates, the intensity of data workers is stable from 2013 to 2015. The intensity of workers is a structural indicator which only changes in the medium to long term.

Moreover, as already explained the data workers are one of the factors of the production function. As the capital increases (ICT), each worker potentially becomes more efficient so that the number of data workers by company does not increase proportionally with the increase in the ICT data technology investments. Finally, it should also be kept in mind that data products and services are at the first stages of their life cycle; at this stage diffusion is not yet pervasive, productivity gains increase rapidly as technology take-up grows, while the number of data workers involved (intensity) does not grow at the same rate.

At the MS level, we confirm that the indicator appears more reliable for the larger countries rather than the smaller ones, particularly the very small countries such as the Baltics, where national specificities in the structure of the economy and the labour market may undermine the general assumptions of the model. Also statistical databases often have gaps for the small MS which we have filled through now-casting methodologies.

Figure 9 Average Number of Data workers by User Company by MS, 2016
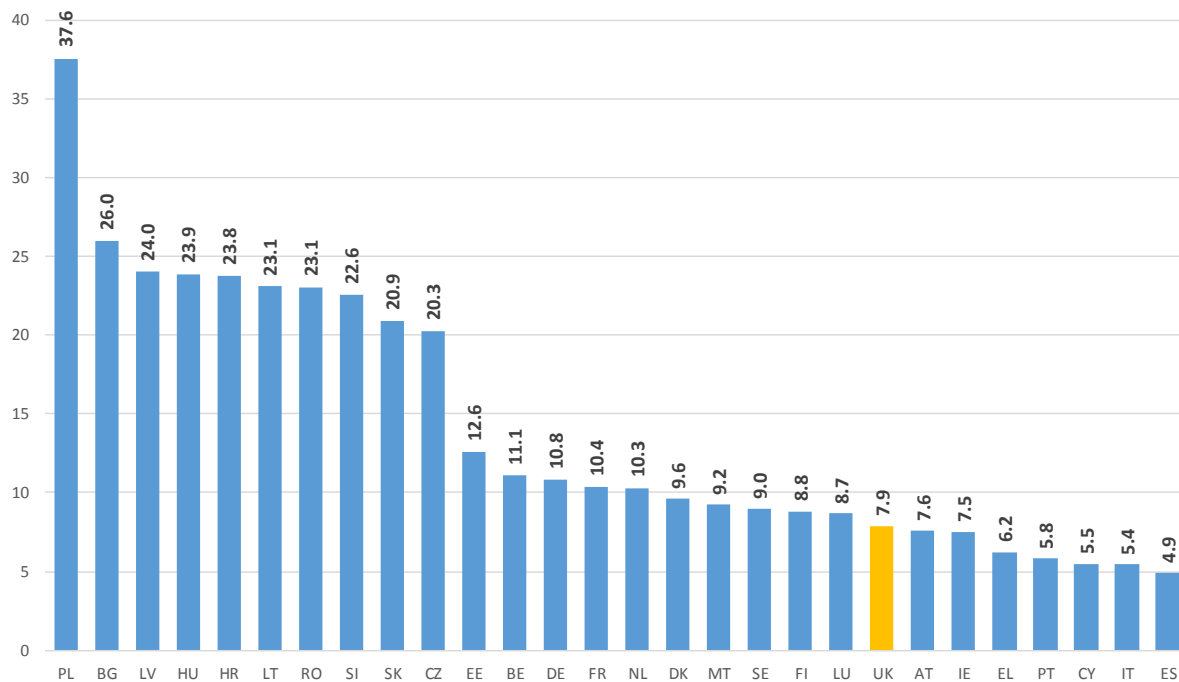
Source: European Data Market Monitoring Tool, IDC October 2016

In some countries the number of workers is definitely above the average: within this group we find Central and Eastern European and Baltic Member all with an average number of data workers ranging from 12 (Estonia) to 37 (Poland). This trend can be explained as follows:

- These countries count a very limited number of data users, in total they represent 4% of the total users in Europe;
- These data users should be the larger companies, and therefore the average number of data workers is higher than in other countries;
- These countries use a lower ICT stock which means that they need more labour to compensate lack of technology.

If we look at their shares of data users on total companies, we can see that they are considerably lower than the rest of Europe: in the Central and Eastern European countries (Estonia, Czech Republic, Lithuania, Latvia, Poland, Slovenia, Croatia, Bulgaria, Slovakia, Romania, and Hungary) in fact only 2% of the companies are on average data users, against the 6% for the rest of Europe. however, the diffusion of data, and therefore of data workers in these countries are limited to the larger and innovative companies.

The more pervasive is a technology, the lower is the intensity share of workers, especially if the country has a high number of SMEs.

In the other countries, the intensity of data workers per user shows a limited variability, from 11% of Belgium to 5% (Spain). Across Europe, 8 countries (Belgium, Netherlands, Germany, France, Austria, Finland, Denmark, and Sweden) present an intensity of data workers per user which is close to the EU average. Some of these countries also account for the highest number of data users, meaning that the use of data and, therefore, of

67

data workers is widespread. Finally, below the EU average we find a small group of countries, including Italy, Spain, Portugal, and Ireland), where the average of data workers per users remains between 8 and 5. In these countries it is well known that there is a large incidence of SMEs.

4.5. Data Workers Forecast

The data workers' indicators are projected to 2020 under the three alternative scenarios, the Baseline, the Challenge and the High Growth. The forecasting approach is based on the idea that countries and industries investing significantly in data products and services will experience an increase in the overall efficiency of labour and capital or in total factor productivity (TFP) growth. Industries receive productivity gains from data use over the labour productivity gains received from investments in ICT, for instance because of an improvement in production processes.

4.6. Indicator 1: Data Worker Forecasts

The number of EU data workers is forecasted to grow in all three scenarios, the Baseline, the Challenge and the High Growth one, as we expect the use of data-driven innovation to increase even in less favourable economic conditions. The forecast indicator represents the total potential demand, therefore it includes a share of potential unfilled positions (which are measured by the data skills gap indicator, see Chapter 10).

Table 13 Forecast Number of Data Workers, 2020, by scenario

Indicator 1 — Data Workers – Forecast 2020								
N.	Region	Name	Description	2020 Challenge (000)	2020 Baseline (000)	2020 High Growth (000)	CAGR Challenge scenario	CAGR Baseline scenario	CAGR High Growth scenario
1.1	EU27	Number of data workers	Total number of data workers in EU (000s)	5,171	6,169	8,134	1.1%	5.7%	13.3%
1.1	EU28	Number of data workers	Total number of data workers in EU (000s)	6,466	7,812	10,431	1.2%	6.1%	14.1%

Source: European Data Market Monitoring Tool, IDC 2016

In the Baseline scenario for EU28, characterized by a healthy growth rate of the data market, the demand for data workers is expected to increase up to 7.8 million, with a compound average growth rate of 6%. In the Challenge scenario, demand will increase at a much slower rate, to about 6.5 million data workers. On the opposite, the High Growth scenario, based on the hypothesis of a demand pull of the data into all the user industries, the demand for data workers is expected to grow to 10.4 million workers with a compound average growth rate of 14%.

SMART 2013/0063 European Data Market

The forecasts by Member State do not radically change the picture emerging from the 2015 data. The distribution of data workers by country remains heavily influenced by the overall size and employment of each country, with the 6 major EU Member States accounting for over 70% of the total number in the three scenarios.

In the Baseline scenario, some Member States show growth rates for data worker demand twice as high as the EU average: they are Croatia, Slovakia, Slovenia, and Netherlands, Malta and Sweden, which start from a lower baseline but are expected to show very positive demand dynamics.

The ranking of MS by demand for data workers does not change much in the Challenge scenario compared to the Baseline scenario: the factors that dampen demand for data workers in the Challenge scenario are expected to reduce the potential for growth in a similar way across Europe compared to the Baseline scenario.

Under the High Growth scenario, ten countries are expected to increase the number of data workers slightly over the EU average. In these countries technology investments will grow faster than in the rest of Europe; they include the Central and Eastern Europe Countries, but also Sweden, UK, the Netherlands, Belgium. The other factor explaining the demand of data workers in the medium to long term is the trend in the total factor productivity[17]. The countries with a less advanced technology may start investing significantly in data technology which explains a faster demand for data workers; countries with an advanced technology may have slow productivity gains (diminishing marginal returns) which may explain as well a high demand of data workers.

In the High Growth scenario, the diffusion of the data technology is faster because the user companies are increasing faster than in the other scenarios. This means that the adoption of the technology and the demand of data workers will depend on the adoption of such technologies by SMEs and by all industries in both advanced and less advanced countries.

---

[17] The Total Factor Productivity (or multi-factor productivity) accounts for effects in total output growth as compared to the growth in the traditional inputs which are labour and capital. As demonstrated by Robert Solow (1957) the TFP is a measure of the technological change since it measures the change of the output not related to the change of the input.

SMART 2013/0063 European Data Market

Figure 10 Data workers forecast by MS, 2020, ranking by number (000s), 3 scenarios

MS	Baseline
UK	1644
DE	1326
FR	821
IT	649
PL	568
NL	477
ES	413
SE	336
BE	208
HU	149
CZ	134
PT	135
RO	132
AT	107
BG	99
DK	88
SK	86
FI	73
EL	70
IE	69
HR	52
LT	44
SI	41
LV	29
EE	26
LU	24
CY	11
MT	11

Legend: Challenge Scenario — High Growth Scenario; Baseline Scenario

Source: Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market
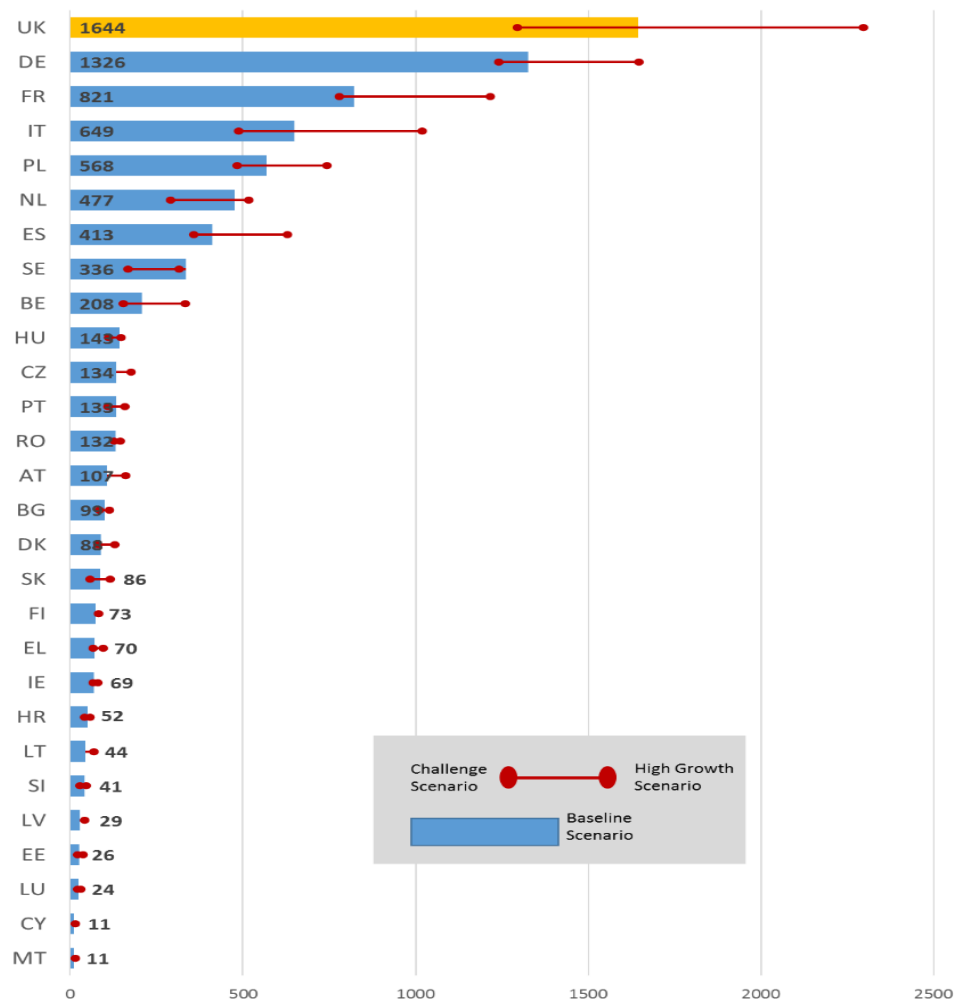
Figure 11 Data workers forecast by MS, 2020, ranking by High Growth scenario (%), 3 scenarios
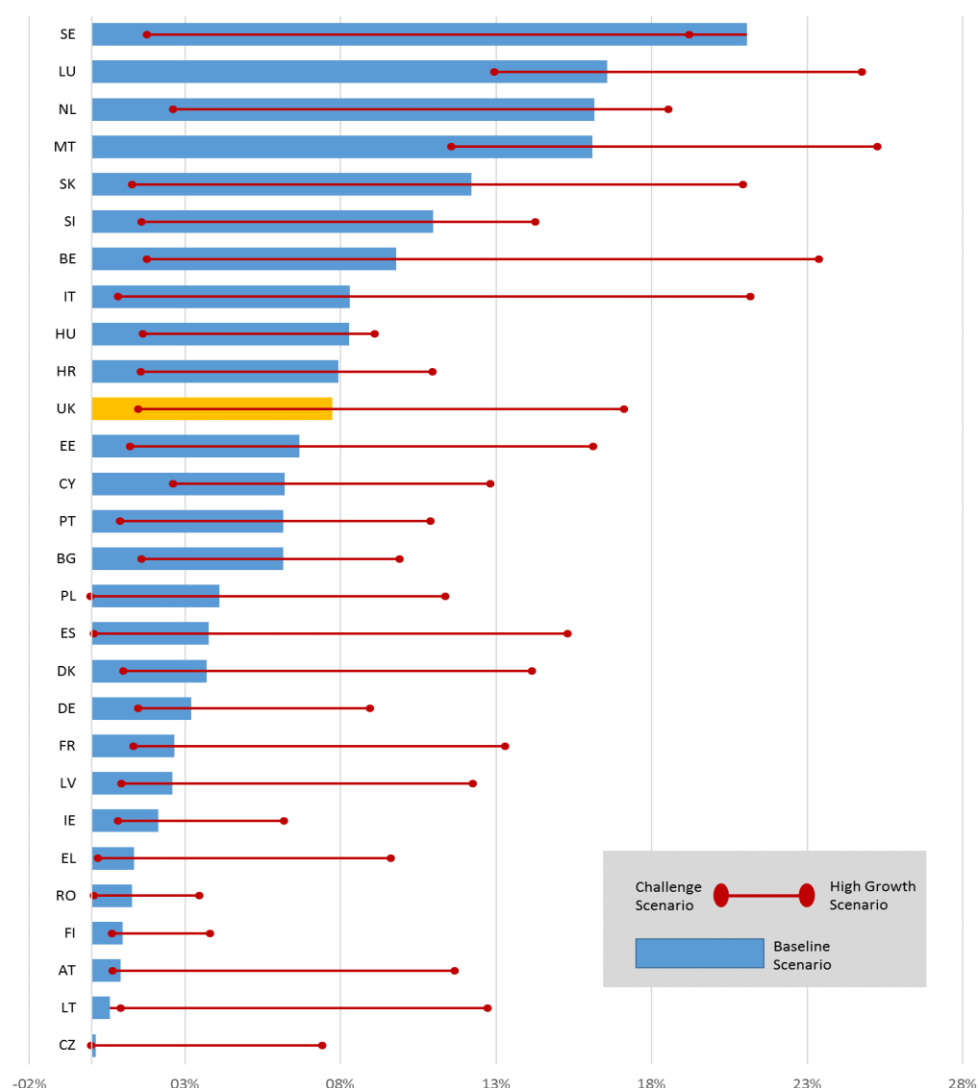
Source: Data Market Monitoring Tool, IDC October 2016

4.7.   Key Findings

These are the key findings from the analysis of data workers.

- 3.1% of total employees in the EU are currently dealing with data produced and managed with digital technology.

Since 2013, approximately 3% of total employees deal with, produce, manage, or use data to improve their decision-making processes. The European Union has currently more than 6.1 million data workers (forecast at 2016). The number of data workers increased by almost 1% year on year in 2014, and then by 3% in 2015 and 2016 year on year.

SMART 2013/0063 European Data Market

- Data workers are required in all countries in the EU, independent of being a "big" or "small" economy.

The penetration of data workers in the labor market is similar in all countries. Although the EU Big Six (France, Germany, Italy, Poland, Spain, the U.K.) account for over 70% of total data workers, the penetration rate of data workers (share of data workers of total employment) in the EU28 is very similar, varying from 2% (Romania) to 6.5% (Luxembourg) with an average penetration rate at 3%. This relates, on one hand, to the ICT penetration rate by country, and on the other hand to the industry structure of countries (relevance of manufacturing versus services industry) and the occupational structure. A high share of managers and professionals makes data necessary for the decision-making processes.

- The growth rate of data workers is correlated with the growth of the number of user companies and therefore is expected to increase strongly in the High Growth scenario

In the Baseline scenario, characterized by a healthy growth rate of the data market, the number of data workers is expected to increase in 2020 to 7.8 million, with a compound average growth rate of 6.0%. In the High Growth scenario, the data workers will reach 10.4 million workers. In the Challenge scenario, demand will increase at a much slower rate, to about 6.5 million data workers.

5 MEASURING THE DATA COMPANIES

5.1 Definition

Data companies *are data suppliers' organizations, whose main activity is the production and delivery of digital data-related products, services, and technologies. They represent the supply side of the data market. On the other hand, 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.*

Therefore, Indicator 2 measures separately:

European data companies, counted as legal entities based in one EU Member States, as a share of the total number of enterprises included in the information and communication industry and professional services industry classification (J and M in the NACE rev2).

European data users, counted as legal entities based in one EU Member State, as a share of the total number of private enterprises in the EU.

As shown in the table below, we counted 661,050 data user companies in the EU28 in 2016 (with a share of 6.4% over the total number of the private enterprises in the EU), up 1.6% against the previous year - the industry comprised 254,850 companies in 2016 (with a share of 14.1% over the total of the information and communication and professional services industries), thus registering a growth year-on-year in units of 2.3% and in share of 1.8%.

In the wake of the Brexit referendum in June 2016, in this Final Report we have also specified the values of Indicator 2 for the EU without the UK (which we have indicated as EU27). The UK being one of the strongest ICT markets in Europe with and a leading data-driven economy, it is not surprising to observe a sharp reduction in the numbers of both data users and data companies in the EU27 vis-à-vis the EU28 across the period under consideration. As a consequence, also the overall share of both data users and data companies in the EU27 will be reduced if compared to the one of the EU28 in the same period.

Table 14 Data Companies and Data Users in the EU, 2013-2014–2015-2016

N.	Name	Description	Market	2013 Value	2014 Value	2015 Value	2016 * Value	Growth 2015/ 2014	Growth 2016/ 2015
2.1	Number of data companies	Total number of data companies in the EU, measured as legal entities based in one EU country	EU27	129,684	128,450	131,900	134,350	2.7%	1.9%
2.1	Number of data companies	Total number of data companies in the EU, measured as legal entities	EU28	239,845	243,600	249,100	254,850	2.3%	2.3%

N.	Name	Description	Market	2013 Value	2014 Value	2015 Value	2016 * Value	Growth 2015/ 2014	Growth 2016/ 2015
		based in one EU country							
2.2	Share of data companies	% share of data companies on total companies in the ICT and Professional services industries	EU27	NA	10.7%	10.8%	10.9%	1.2%	1.0%
2.2	Share of data companies	% share of data companies on total companies in the ICT and Professional services industries	EU28	13.8%	13.6%	13.8%	14.1%	1.7%	1.8%
2.3	Number of data user companies	Total number of data users in the EU, measured as legal entities based in one EU country	EU27	487,355	494,400	499,650	506,400	1.4%	1.1%
2.3	Number of data user companies	Total number of data users in the EU, measured as legal entities based in one EU country	EU28	633,605	642,700	650,750	661,050	1.3%	1.6%
2.4	Share of data users	% share of data users on total companies in the EU industry	EU27	NA	5.6%	5.6%	5.7%	0.5%	0.8%
2.4	Share of data users	% share of data users on total companies in the EU industry	EU28	6.2%	6.3%	6.3%	6.4%	0.7%	1.1%

*2016 Estimates

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market

Indicator 2 — Data Companies								
N.	Name	Region	2020 Challenge (000)	2020 Baseline (000)	2020 High Growth (000)	CAGR Challenge scenario	CAGR Baseline scenario	CAGR High Growth scenario
2.1	Number of data companies	EU27	139,500	163,550	188,000	0.9%	5.0%	8.8%
2.1	Number of data companies	EU28	265,250	310,250	359,050	1.0%	5.0%	8.9%
2.2	Share of data companies	EU27	11.3%	13.3%	15.3%			
2.2	Share of data companies	EU28	14.7%	17.1%	19.8%			
2.3	Number of data user companies	EU27	511,250	556,850	820,650	0.2%	2.4%	12.8%
2.3	Number of data user companies	EU28	668,400	727,250	1,098,600	0.3%	2.4%	13.5%
2.4	Share of data users	EU27	5.7%	6.2%	9.2%			
2.4	Share of data users	EU28	6.5%	7.1%	12.8%			

*2016 Estimates

Source: European Data Market Monitoring Tool, IDC October 2016

The majority of data companies are also data users, but not all of them are such by definition.

The overlapping between demand and supply actors is reflected by the Figure below, which also outlines the different size of the two stakeholder populations. The number of data companies is much smaller than the number of data users, but is also more concentrated, coming from only 2 sectors. Consequently, data companies represent a relatively high share of the ICT and professional services sectors, with strong growth dynamics (as we will see in the forecast scenarios).

The data users come from all sectors of the economy, and represented a much lower share of the European private companies' population (between 6.3% and 6.4% across 2014, 2015 and 2016), hence with a strong growth potential over the coming years.

Figure 12 Data Companies and Data User Companies in the EU (2015 Update)

SMART 2013/0063 European Data Market
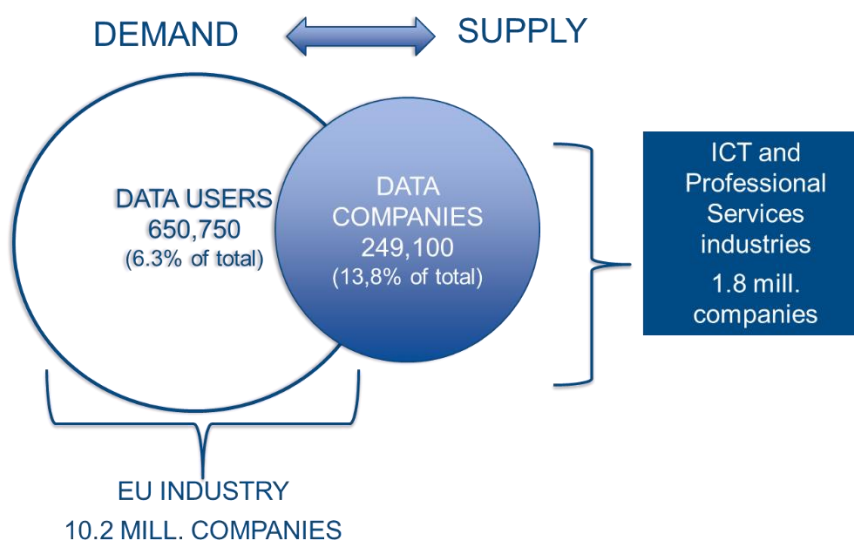
Source: European Data Market Monitoring Tool, IDC 2016

5.1.1 Updating the Measurement of Data Companies in 2016

Indicators 2.1, 2.2, 2.3 and 2.4 were updated in in view of this Final Report by leveraging a series of internal IDC sources, as well as public sources from EU and national statistical offices. Among them, the following are worth mentioning:

- 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: June 2016);
- Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: June 2016);
- IDC's detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 2015 and 2016;
- IDC Worldwide Black Book (Standard Edition), quarterly updates form the years 2015 through 2016 (latest available figures for the second quarter of 2016). 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
- IDC European Vertical Markets Survey, 2015: More Western European SMBs Will Invest in Software Solutions Beyond Maintenance, July 2015
- IDC Big Data Use Cases and Future Developments in European Vertical Markets in 2016: An IDC Survey, November 2016
- IDC Western European SMBs' Mobility Landscape, 2016: An IDC Survey, September 2016

To fully apprehend the effects of the results of the Brexit referendum on the future development of Indicator 2, the study team also leveraged IDC's ad-hoc research summarized in the document below:

- The Brexit Impact on IT Spend in the U.K. and Western Europe: A Scenario Analysis : An IDC Insights, July 2016.

SMART 2013/0063 European Data Market

For Data Suppliers the number of companies counted were limited to those in the industries associated with the data market – specifically those in the ICT and Professional Services industries, as defined by the NACE-R2 codes in the methodology. The definition for the source data for the number of Data User companies is broader than that of Data Suppliers, encompassing a much wider range of industry segments as defined by the NACE-R2 codes in the methodology document.

5.1.2 Data Companies

Data companies may be start-ups, innovative SMEs, or spin-offs of large enterprises that are engaged in developing and exploiting the new Big Data technologies in innovative ways. At the same time, the data market is also attracting many ICT, business analytics, or information services enterprises that are renewing their offerings to develop data-driven products and services. The data industry therefore includes both "pure players" whose core business is data-driven products and services, and "mixed players" that combine traditional and data-driven offerings. Figure 13 shows how we have classified data companies based on our definition of the data market and the taxonomy developed for this study.

Figure 13 Classification of Data Companies
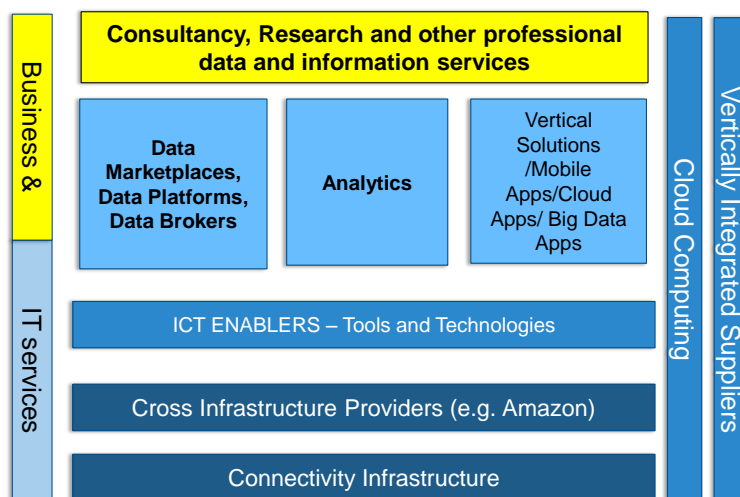
Source: IDC 2014

Legend: Blue and light blue boxes include enterprises classified in the Information and Communication sector (J); Yellow boxes include enterprises classified in the Professional Services sector (M).

The figure includes enterprises classified in the information and communication sector (blue and light blue boxes) and enterprises classified in the professional services sector (yellow boxes). They can be described as follows:

- **Providers of consultancy, research, and other professional data and information services.** They are, for example, the publishers of online directories, credit information, and market research companies exploiting Big Data related tools and technologies to update their offering. They integrate the many phases of the data value chain, from data creation and collection, to storage, analysis, primary and secondary use. For the sake of this study we have taken into account only the activities, revenues, and employees related to the provision of data-based services.

- **Providers of business and IT services.** This includes business consulting, business process outsourcing, IT project-based services, network consulting and integration services, IT outsourcing, storage services, security services, software and hardware support, and training services related to Big Data implementations.
- **Enterprises leveraging data technologies** to help other organisations (data holders and users, who sometimes are the same) to exploit and use data in innovative ways. They are the pure players of the data market, which is their core business. They play the role of specialised intermediaries between data holders and data users and many of them are start-ups or innovative SMEs. Their offering can be classified as follows:
  - **Providers of data marketplaces and data platforms** where data is stored, curated, and exchanged. This includes marketplaces such as Datamarket.com and repositories such as Knoema. They provide a mix of SaaS services and data, premium datasets, and access to complex and diverse data sources.
  - **Providers of analytics and discovery software**, including search engines, data mining, text mining and other text analytics, rich media analysis, and data visualization.
  - **Providers of vertical solutions/mobile apps/cloud apps/Big Data apps:** application software including business process or industry-specific applications such as for web clickstream analysis, fraud detection, and logistics optimization.
- **ICT enablers.** They provide tools and services enabling the management, storage, processing, analysis, and distribution of data. They provide for example information management software, including parallel and distributed file systems with global namespace, highly scalable (size and structure) relational databases, key value pair (KVP) data stores, content management systems, graph databases, XML databases, object-oriented databases, dynamic application data stores and caches, data integration, and event-driven middleware.
  - ICT infrastructure providers including:
    - Cloud computing providers
    - Providers of platforms and IT infrastructure
    - Connectivity infrastructure providers

Finally, one more group of stakeholders needs to be highlighted:

- **Vertically integrated suppliers:** large organisations that leverage their own data to create a new business in the provision of specific data-related business (telecom operators, utilities, financial services), or OTT (over the top) global players such as Google, Facebook, Microsoft, and IBM.

For this study, for these categories of suppliers we have taken into account only the activities, revenues, and employees related to the provision of data-based services.

5.1.3 Data User Companies

The demand side of the data market is represented potentially by all enterprises, since every organisation uses data. There are some industries which are more intensive users of data. Finance, healthcare, and retail (particularly ecommerce) are industries where data has a significant and strategic role in the decision-making process; most companies have important datasets, and use them and in some cases sell/exchange them. Data users therefore include all the economic sectors, aggregated into the eleven main industries used for this study.

5.1.2.1 Data Users Who Are Also Data Suppliers

The boundaries between demand and supply are not so clear-cut, since the companies which develop a good capability to exploit their own data may become in turn resellers of their own data to third parties. This is particularly true for enterprises active in the B2C market that increasingly monitor their customers' activities collecting data, for example from the retail, finance and even education industries. For example, the U.K. retail giant Tesco (according to media news) is engaged in leveraging the information on the spending habits of its own customers, including the 16 million members of its Club card loyalty scheme. On the other hand, these activities tend to generate protests by consumer organisations and not all organisations are willing to deal with potential data and privacy protection issues. However, since in many industries the potential of the exploitation of one's own data is very high, it is likely that in the future companies from the retail, finance, and other services industries may become data companies in their own right (or they may launch spin-offs). However, at present these companies are more the exception than the rule and we have not included them in our data companies' indicator.

From the point of view of the measurement of our indicator therefore we have decided the following:

- Traditional companies with a division or business unit dedicated to the development of data products and services are not data companies; if the division or business unit becomes a separate company (a spin-off) then it becomes a data company in its own right and is included in our definition of the data industry.
- At this stage, in the main B2C sectors such as finance, retail, and durable consumer goods (automotive) there are still barriers to the resale of customer data to third parties; to our best estimate, the spin-off of data companies from traditional companies is still quite rate.
- In addition, it would be misleading to include in the estimate of the data market the full revenues for example of Tesco, only because it has a marginal activity in data reselling.

5.2   Indicator 2.1: Data Companies

5.2.1 Data Companies by Member State

All in all, in 2016 the data industry in Europe featured almost 255,000 companies (up 2.3% from the previous year) representing a share of 14.1% of enterprises populating the ICT and professional services sectors in the same year amounting to slightly more than 1.8 million companies - as a reference, data companies represented a share of 13.8% over a total of 1.8 million ICT and professional services in 2015. The above results confirm that European companies continue their journey towards innovation through the adoption of data-related technologies and the subsequent development of data-based products and services.

As in the previous years, the number of data companies by Member State in the EU in 2016 continued to be heavily concentrated in the two major countries, the U.K. and Germany. The distribution of data companies reflects the concentration by country of the ICT industry.

Figure 13 displays both the distribution of data companies by Member State as well as their growth rate year-on-year. The landscape is dominated by the U.K., followed by the largest EU member states, with a long tail of small groups of data companies in most countries. This distribution is not simply a mirror of the size of each of the member states economies, but is more closely correlated with the presence of a strong ICT industry, and a dynamic professional services industry.

Figure 14 Ranking of Member States by number of Data Companies, 2016, 000s; Data Companies Growth 2016-2015 by Member State, %
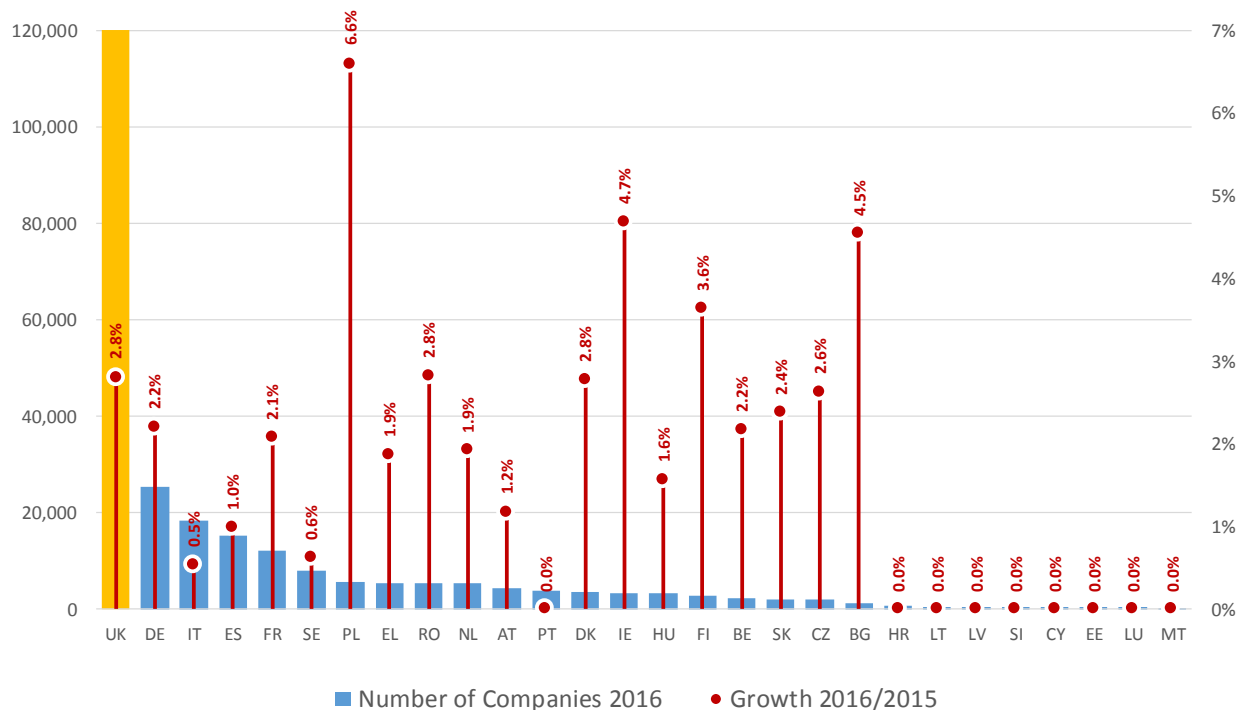
Source: European Data Market Monitoring Tool, IDC October 2016

The growth rate for the number of data companies between 2016 and 2015 is estimated at 2.3% - in line with the growth rate of the period 2015-2014, thus showing a moderate consolidation of the data-driven economy in the EU as a whole.

Another way of representing the development stage of the data industry by Member State is to display the intensity of data companies' presence measured as the share of data companies on the total number of companies in the reference sectors by country.

SMART 2013/0063 European Data Market

Figure 15 Number of Data Companies by MS (in '000) and Share of data companies by MS of total J and M sectors, % 2016
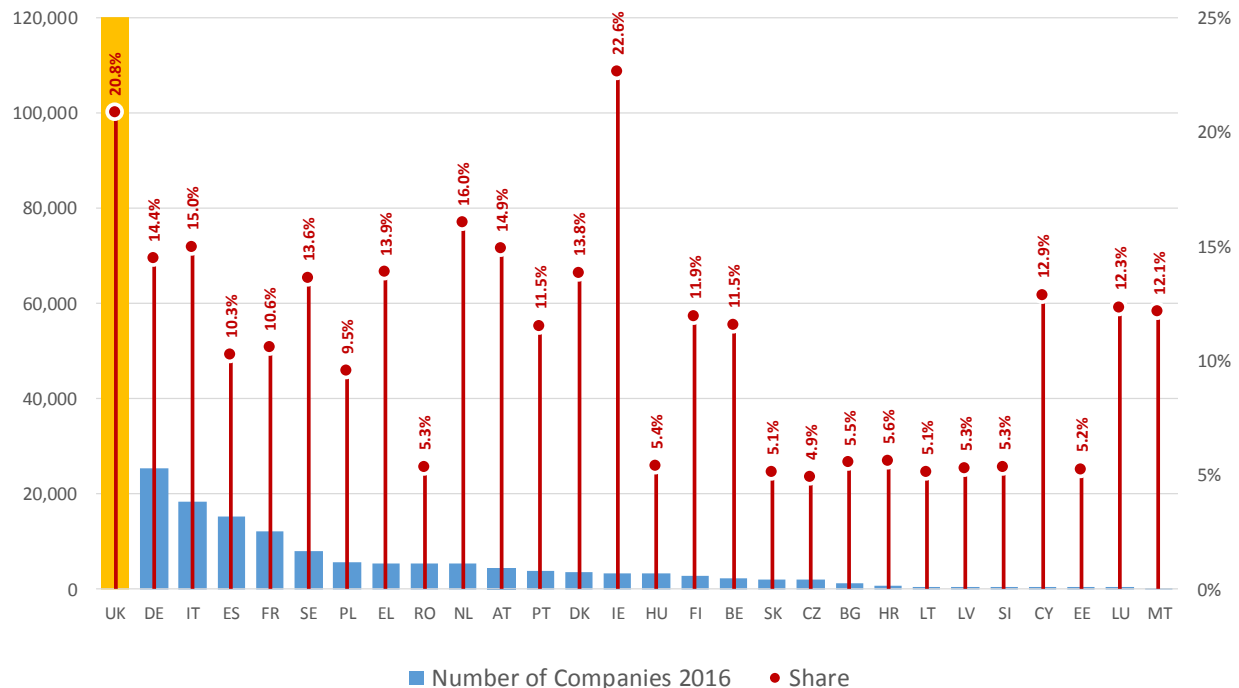
Source: European Data Market Monitoring Tool, IDC October 2016

With an average share of 14.1% at EU28, we can distinguish between four main groups of countries with similar shares of data companies' penetration:

- A restricted group of unchallenged front-runner made up of the **U.K. and Ireland**, with shares around or over 20%. Ireland pursues its promotion programme focused on Big Data and Big Data research and the UK government continues to foster the development of a local and multinational data industry on its territory.
- A limited group of second-best countries with **the Netherlands, Italy, Austria, Germany, Sweden, Denmark and Greece** features a medium-high level of penetration with a share slightly above or aligned with the EU28 average. All of these countries, including Greece until recently, used to have a lively ICT industry and it is no surprise that they can all boost a relatively high presence of data companies in both 2014 and 2015.
- A third group of Member States are below the EU average, including **Cyprus**, **Luxembourg, Malta, Belgium, Finland, Portugal, France, Spain** and **Poland**. France and Spain have a sizable number of data companies but a large proportion of small, traditional companies in the J and M sectors that may take time before they take up data technologies. Poland has good potential but, again, a still traditional industry base.
- The last group includes all the Central and Eastern European Member States with shares of data companies around or slightly below 5% - much lower than the EU average: **Croatia, Bulgaria, Hungary, Slovenia, Romania, Latvia, Estonia, Lithuania, Slovakia and the Czech Republic.** Central and Eastern European Member States may still be considered as a potential area of concern in terms of data industry growth and development. Some of these Member States keep exhibiting growth rates higher than the EU28 average in 2016 – **this is the case of the Czech**

81

SMART 2013/0063 European Data Market

Republic, Slovakia and Romania – thus showing that the data industry is indeed redeveloping in these countries, although with a certain delay if compared to most of the Member States in Western Europe.

5.2.2 Data Companies by Industry

In 2016, 34.1% of the EU28 data companies came from the ICT industry (they were 33.4% in 2015) and 7.6% were issued from the professional services industry (as opposed to 7.5% the year before). Although extremely moderate, both industries are witnessing an increase in penetration in terms of data companies, with the ICT sector growing faster than the professional services sector, which is logical given the traditional proximity of the ICT sector with data-related technologies.

Table 15 Number and share of EU Data Companies by Industry, 2013-2014-2015-2016

	Industry	Information and communications	Professional services	Grand Total
Number of data companies, Units	2013	153,425	86,420	239,845
	2014	143,250	100,350	243,600
	2015	146,650	102,400	249,100
	2016*	150,500	104,350	254,850
Growth rate, %	2015/2014	2.4%	2.0%	2.3%
	2016/2015	2.6%	1.9%	2.3%
Share of total potential data companies, %	2014	32.7%	7.4%	13.6%
	2015	33.4%	7.5%	13.8%
	2016*	34.1%	7.6%	14.1%
Total number of potential data companies, Units	2013	420,300	1,312,500	1,732,800
	2014	437,750	1,353,800	1,791,550
	2015	439,300	1,361,400	1,800,700
	2016*	440,900	1,368,950	1,809,850

*2016 Estimates

Legend: Potential data companies belong to the sectors J – Information and Communication and M – Professional services

Source: European Data Market Monitoring Tool, IDC October 2016

5.2.3 Data Companies by Company Size

As in the previous round of measurement of the European Data Market monitoring tool, the SMEs' share of the EU data industry is estimated to be approximately 98.9% in 2016, a bit lower than the SMEs share in the reference sectors which is 99.5%. This reflects the structure of the EU industry, for example, the fact that there are several hundreds of thousands of small IT companies reselling IT products and services, who are entering the data industry to answer demand. However, medium and large companies do play an important role in the industry, and this is reflected by the data when looking at the penetration rates of data companies by two size classes. In particular:

- The 1-249 employees' size class detains a share of total of 14% of the total potential data companies in 2016 (up of a tiny 0.3% with respect to 2015);
- The 250+ employees' size class features more than 2,800 companies in 2016 totalling a 32.1% of share on the total number of potential data companies (again slightly on the increase – it was 31.3% in 2015).

Table 16 Number and share of Data Companies by company size, 2013-2014-2015-2016

	Size Band	1-249 empl.	250+ empl.	Total EU28	SMEs share
Number of data companies, Units	2013	237,128	2,717	239,845	98.9%
	2014	240,840	2,760	243,600	98.9%
	2015	246,313	2,787	249,100	98.9%
	2016*	252,000	2,850	254,850	98.9%
Growth rate, %	2015/2014	2.3%	1.0%	2.3%	
	2016/2015	2.3%	2.3%	2.3%	
Share of total potential data companies, %	2014	13.7%	33.4%	13.8%	
	2015	13.7%	31.3%	13.8%	
	2016*	14.0%	32.1%	14.1%	
Total number of potential data companies, Units	2013	1,724,663	8,137	1,732,800	99.5%
	2014	1,783,088	8,412	1,791,500	99.5%
	2015	1,791,800	8,900	1,800,700	99.5%
	2016*	1,800,975	8,875	1,809,800	99.5%

*2016 Estimates

Legend: Potential data companies belong to the sectors J and M services

Source: European Data Market Monitoring Tool, IDC October 2016

5.3 Indicators 2.1: Data Companies Forecast

5.3.1 Data Companies Forecasts

The number of data companies at 2020 varies substantially by scenario, as shown by figures 15, 16 and 17 below.

According to our most recent estimates, data companies are expected to grow at a compound annual growth rate (CAGR) of 5% from the period 2016-2020 under the Baseline scenario growing to a total of more than 310,000 units in 2020 – a very minor improvement with respect to our previous estimate at 2020 under the Baseline scenario.

In the Challenge scenario the CAGR from 2016 to 2020 is expected to be only 1% leading to a total number of data companies' units in 2020 of almost 266,000.

According to the High Growth scenario we reckon that data companies will amount to 359,050, growing at CAGR of 8.9% over the period 2016-2020 – a slight decrease in terms of absolute numbers but at a faster-paced growth with respect to our previous estimates for the year 2020 in the previous round of measurements.

Figure 15 displays the number of data companies by Member State in the three scenarios, ranked by size.

The UK, Germany, Italy, Spain and France features therefore in the highest upper part of the ranks with the highest number of data companies under all three scenarios.

The figures below present the updated results for the total EU at 2020 under the three scenarios in consideration indicating that data user companies too exhibit a solid growth path in the period 2015-2016 and throughout the year 2020.

Figure 16 Forecast of Data Companies Number by Member State, 2020, 000s, three scenarios
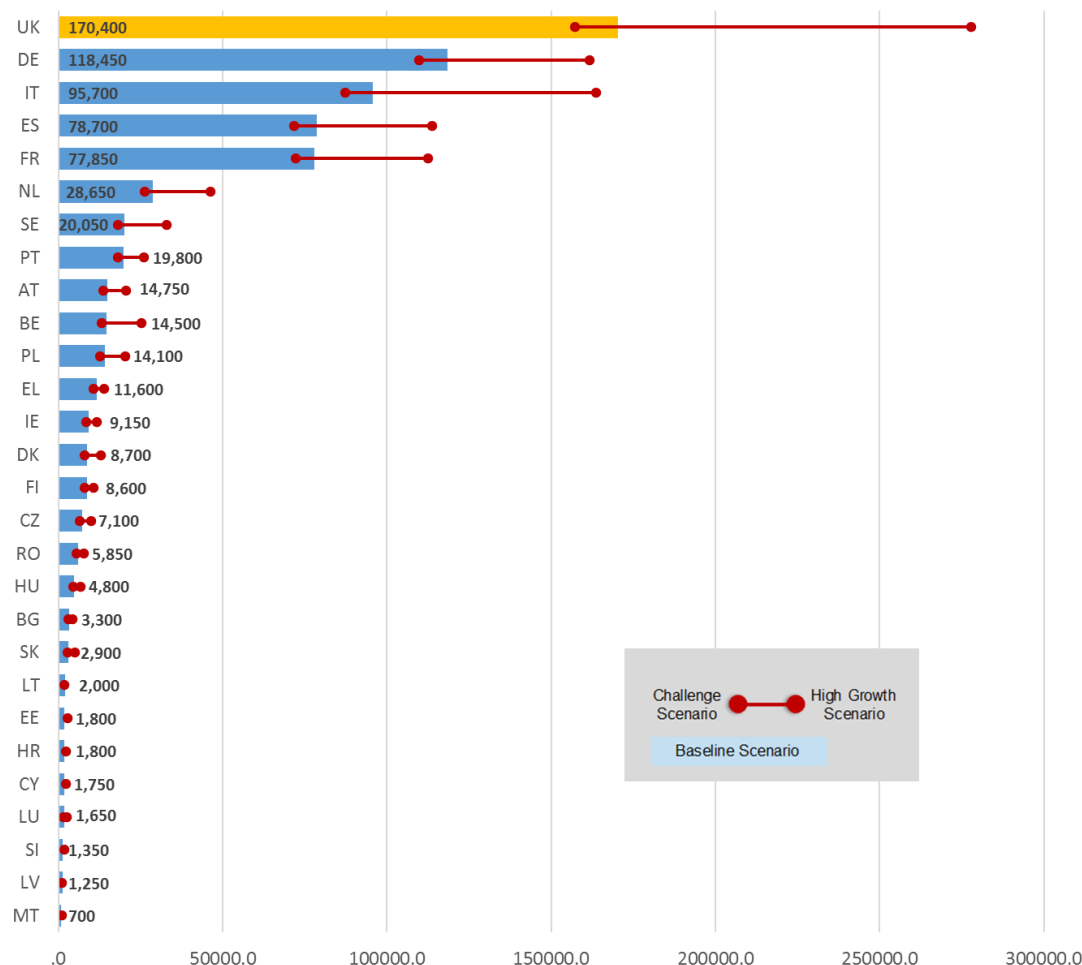
Source: European Data Market Monitoring Tool, IDC October 2016

In terms of growth rate, we anticipate the number of data companies to increase in the EU from the year 2016 to the year 2020 at a CAGR of 5% under the Baseline scenario, which is going to plummet at 1% in the Challenge scenario and increase to 8.9% according to the High-Growth scenario. In the context of a Baseline scenario, the smaller and relatively well ICT-penetrated EU economies will perform better, with growth rates beyond the EU average. This is notably the case of Sweden, Belgium and The Netherlands but also of some smaller economies such as the Baltic countries, Slovenia and Croatia.

Under the Challenge scenario, the EU will see its number of data companies growing only by a moderate 1% in the period 2015-2020. Again Luxembourg, The Netherlands, Sweden, Belgium, Slovenia and Croatia will exhibit growth rates above the EU28 average.

In case the next five years are characterized by both moderate economic growth and a significant impulse of ICT investments (High Growth Scenario), Denmark will rank first, followed by Ireland, Finland, Slovenia and Belgium.

SMART 2013/0063 European Data Market

Figure 17 Data companies CAGR 2020/2016 by Member State
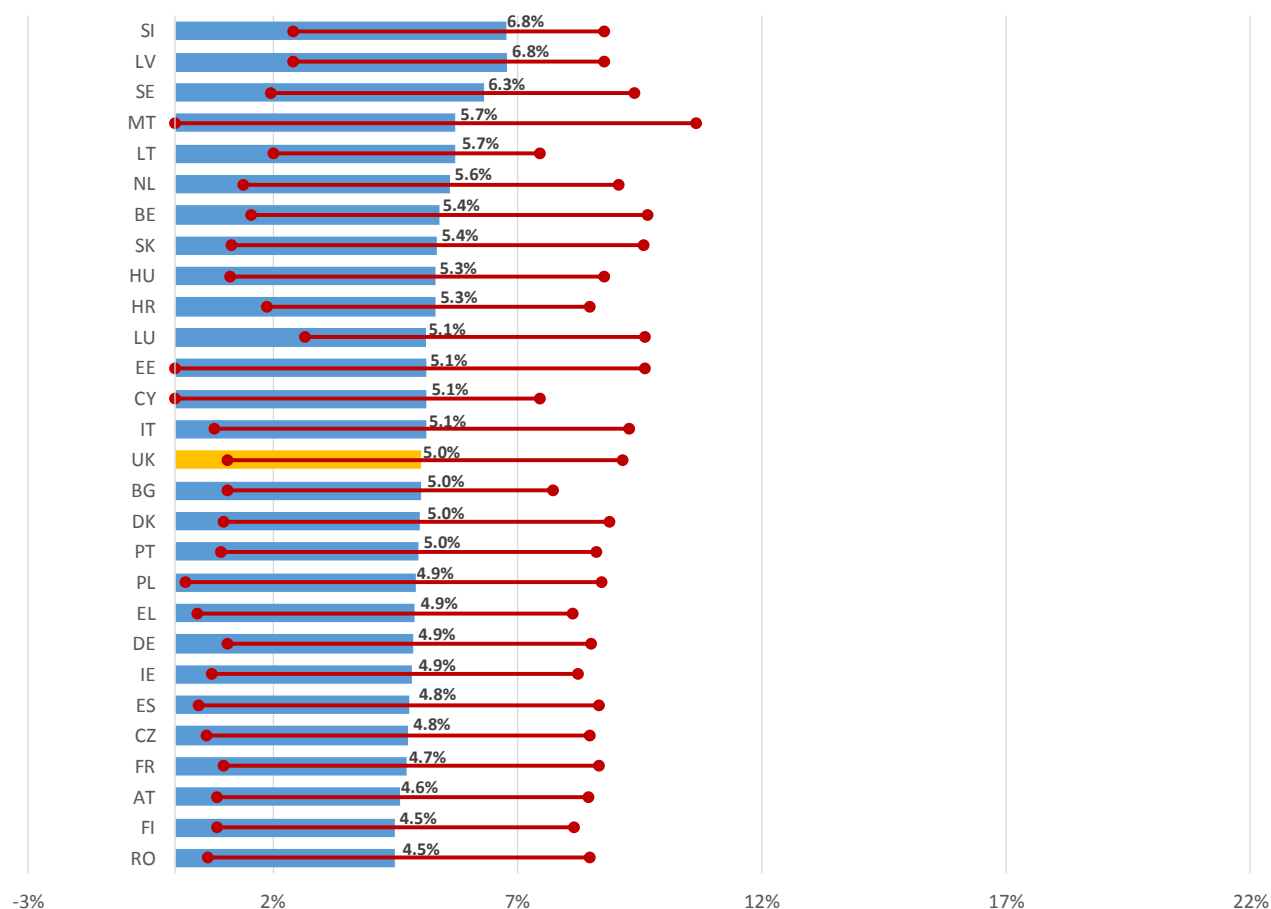
Source: European Data Market Monitoring Tool, IDC October 2016

Figure 17 shows the forecast share of data companies on total by Member State. The ranking by share confirms the main clusters of MS identified in 2016, with limited variations by scenario.

- **The U.K. and Ireland** come at the top under all three scenarios nearing or going beyond the 29% and 31% share in the High Growth scenario and reaching more than 25% and 27% respectively in the Baseline scenario.
- **The Netherlands, Italy, Austria, Denmark and Sweden** present shares of data companies above the EU average, that is between 17.4% and 27.4% in the Baseline scenario, 14.7% and 23.3% in the Challenge scenario and 19.5% to 31.1% in the High Growth scenario.
- **Cyprus, Malta, Luxembourg, Finland, Belgium and Portugal** display data companies' shares aligned or slightly below the EU28 average;
- All the remaining Member States, and in particular the countries in **Central and Eastern Europe,** as well as the **Baltic states** feature value shares well below the EU28 average in all three scenarios.

SMART 2013/0063 European Data Market

Figure 18 Data Companies forecast share of total companies, by Member State
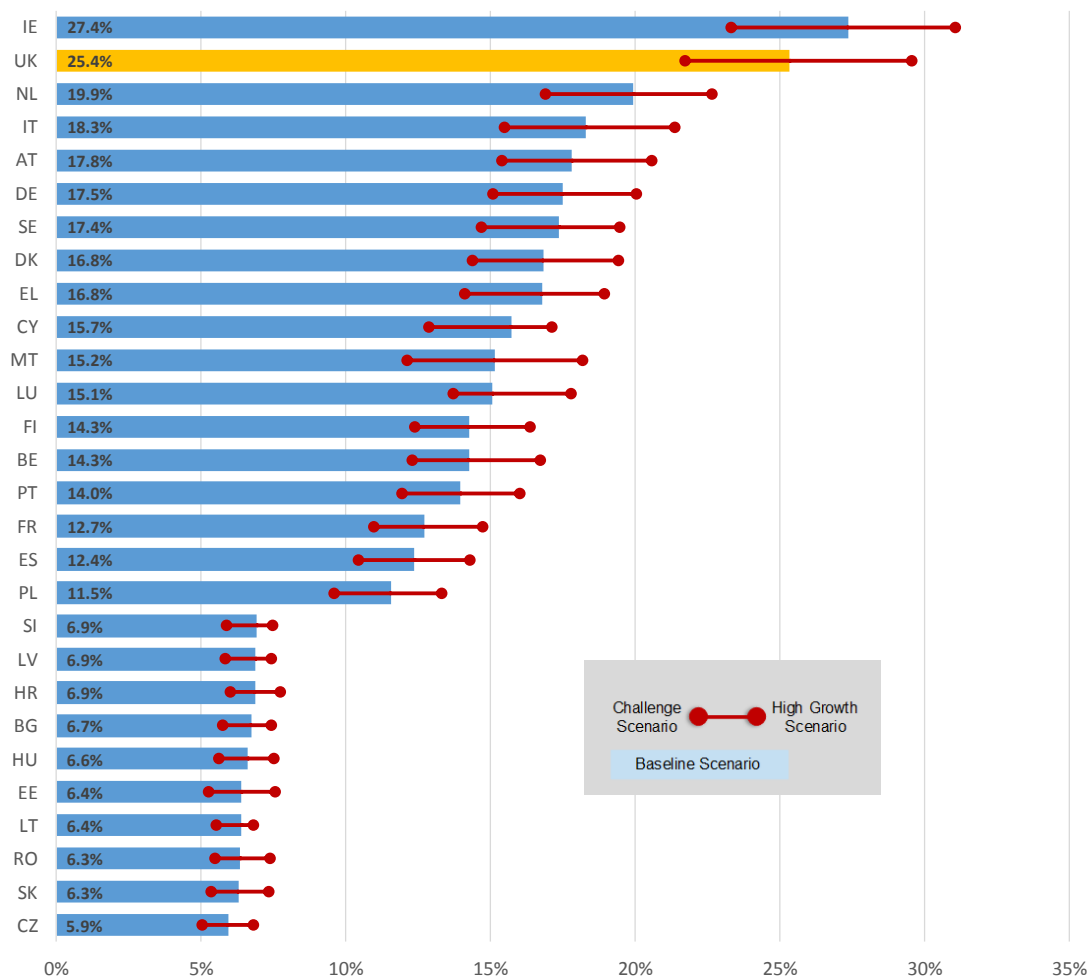
Source: European Data Market Monitoring Tool, IDC October 2016

5.3.2 Data companies forecast by size class

As for the analysis relative to the period 2015-2016, we also present the data companies forecast to the year 2020 under the three scenarios by two main size classes according to the number of employees.

We notice that companies tend to follow different trends according to their size in employees. Small and Medium companies (1-249 employees) are more numerous and are expected to grow at higher rates than companies employing 250 or more employees. Our analysis further reveals that the difference in size has little bearing when negative assumptions are made about both the overall economic growth and the pace of ICT investments over the next four years (Challenge scenario). On the contrary, when moderate or sustained growth is introduced in the picture, larger companies behave significantly differently from small and medium companies as in the case of the Baseline and Challenge Scenarios.

SMART 2013/0063 European Data Market

Table 17 below summarizes our most recent results in terms of data companies forecast by size band for the total of the EU under the three scenarios under consideration. Table 18 offers further details in terms of share of data companies on the total companies in the period 2014 – 2016 and according to the three forecast scenarios.

Table 17 Forecast of Data Companies Number by company size, 2020, three scenarios

	Number of Data Companies, Units			CAGR 2020/2016 of Data Companies, %		
Size Band	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario	Challenge Scenario	Baseline Scenario	High Growth Scenario
1-249 empl.	262,200	306,800	355,000	5.0%	4.5%	8.9%
500+ empl.	3,050	3,450	4,100	1.0%	4.4%	9.0%
Total EU28	265,250	310,250	359,100	1.0%	5.0%	8.9%
SMEs share	98.9%	98.9%	98.9%			

Source: European Data Market Monitoring Tool, IDC October 2016

Table 18 Share of data companies on total potential companies, forecast by size class by scenario

Share of data companies of total potential companies, %						
Size Band	2014*	2015	2016*	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
1-249 empl.	13.7%	13.7%	14.0%	14.6%	17.0%	19.7%
250 + empl.	33.4%	31.3%	32.1%	34.4%	38.9%	46.2%
EU28	13.8%	13.8%	14.1%	14.7%	17.1%	19.8%

* 2016 estimate

Source: European Data Market Monitoring Tool, IDC 2016

5.3.3 Key findings

- The data industry as a whole comprised approximately 255,000 companies in 2016 in the total of the EU, representing 14.1% of the 1.8 million enterprises populating the ICT and professional services sectors in the same year. The European Data Market monitoring tool confirms a constant and smooth growth in the number of data companies in the period 2013-2016 – the industry counted more than 249,000 companies in 2015, 243,000 companies in 2014 and almost 240,000 in 2013.
- The industry is dynamic and continues to attract start-ups, innovative SMEs, as well as a growing number of more traditional and existing enterprises that make a more intensive use of data and data-related technologies for business purposes.

SMART 2013/0063 European Data Market

- The overall number of data companies grew 2.3% year-on-year in both 2015 and 2016 with data companies pertaining to the ICT sector being slightly more active (i.e. growing more rapidly) than those belonging to the professional services industry.
- A very high concentration rate emerges when looking at the distribution of data companies by Member State: as a matter of fact, the UK, Germany and Italy muster almost two thirds of all data companies in the EU in the period 2013 - 2016 and the picture does not significantly change when forecasting our results to the year 2020. This distribution reflects the relative strength of the ICT industry in general and of the data market industry in particular.
- The majority of EU data companies come from the ICT industry, with a 34.1% penetration rate in total. The other data companies come from the professional services industry, representing a share of 7.6% of the total. When compared to the results presented in our previous report, these percentages demonstrate a certain dynamism from the side the professional services' industry to the detriment of the ICT industry. As the data industry matures there are likely to be more suppliers coming from other sectors as well (finance, retail) that will focus more on the exploitation of data rather than on data technologies per se.
- The SME share of the EU data industry is stable at 98.9% in 2013, 2014, 2015 and 2016, corresponding to a penetration rate of 14% in 2016, while for enterprises over 250 employees it is 32.1%.
- We estimate the number of data companies in 2020 to increase considerably, especially under the Baseline and High Growth scenarios. Under the Challenge scenario, data companies in the total EU could amount to more than 259,000 in 2020 marking a Compound Annual Growth Rate (CAGR) of 1%. The same CAGR could be at 5% under the Baseline scenario and would reach 8.9% in the High Growth scenario.
- The exit of the UK from the EU will not substantially change the picture but it will exert a somewhat negative influence, especially in the short to medium term. The number of data companies in the EU27 (EU28 less the UK) will grow of 1.9% year-on-year in 2016 as opposed to 2.3% in the same period for the EU28. In 2016 the overall share of data companies on the total of the industry will be lowered of more than 3 percentage points if the UK were not to be included in the calculation (total share of data companies in the in the EU27 would amount to 10.9% in 2016 as opposed to 14.1% in the EU28 in the same year).
- The impact of Brexit should be smoother in 2020 with only very minor variations in the CAGR 2016-2020 under the three scenarios considered – the Baseline scenario should remain unaffected (with a CAGR of 5% over the period in both the EU27 and EU 28; the Challenge and High Growth scenarios would only suffer a minimal 0.1% with a CAGR 2016-2020 of 0.9% under the Challenge scenario in the EU27 vis-à-vis a CAGR of 1% in the EU28 and a CAGR of 8.8% in the EU27 vis-à-vis a CAGR of 8.9% in the EU28).

5.4 Indicator 2.2: Data Users

5.4.1 Data Users by Member State

In 2016, data users in the EU28 will reach more than 661,000 units according to our latest estimate, thus growing 1.6% year-on-year and representing an overall penetration of 6.4% over the 10.3 million potential user companies – a persisting low penetration across the four years under consideration (6.2% in 2013 and 6.3% in 2014 and 2015).

As opposed to data companies, data users can be found in every industry sector and are therefore less concentrated than the former, also in terms of Member State distribution. Still, the UK leads the way in this respect too, followed by Germany, Italy and Spain, which muster alone almost two thirds of the total number of the European data user companies. Figure 18 also reveals that some countries continue to exhibit growth rates in 2015 that are way above the EU average: this is the case of the Netherlands, the UK, Sweden, Cyprus, the Czech Republic and, to a lesser extent, France and Germany, that confirm the positive trend already displayed in 2015 over the previous year and show above-average year-on-year growth rates in 2016.

Figure 19 Ranking of Member States by Data Users population, 2015-2016, 000s
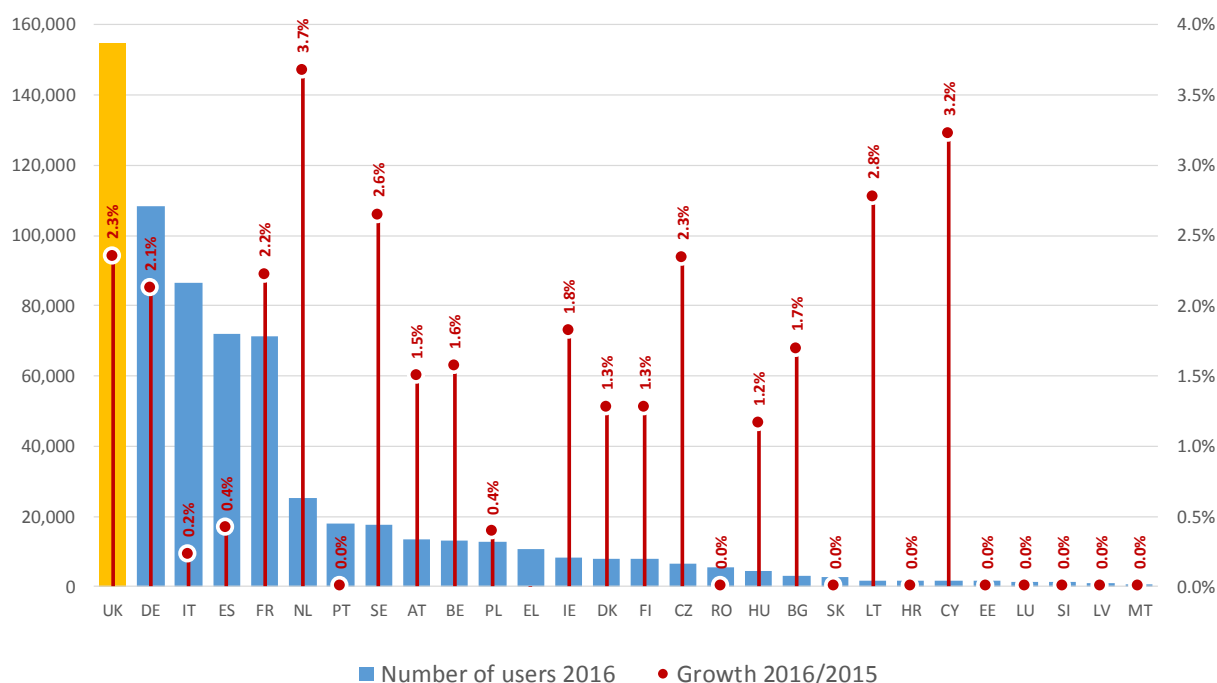
Source: European Data Market Monitoring Tool, IDC October 2016

To complete the analysis of data users by Member State, we have looked at the share of users on total companies as in the previous report (Figure 19). Not surprisingly, the differences by Member State are less pronounced than in the case of data companies as user companies constitute a larger group than supplier companies and mirror more closely the overall structure of the economy. Nonetheless, it is possible to distinguish at least three main groups with similar characteristics:

SMART 2013/0063 European Data Market

- The **Netherlands**, the **UK, Ireland** and **Luxembourg** emerge as clear leaders with a percentage share of user companies between 9% and 12% of their total amount of companies.
- A composite group made of larger economies like **France, Germany,** and **Italy**, complemented by smaller Member States like **Sweden, Austria, Finland, Portugal,** and **Denmark** exhibits data users' shares between 6.4% (the EU average) and 9%.
- Spain, Cyprus, Estonia, the Czech Republic, Lithuania, Latvia, Poland and Slovenia lie between 2% and the EU28 average.
- Finally **, Greece, Hungary, Bulgaria, Slovakia, Croatia** and **Romania** close the rank with a share of data user companies between 1% and 2%, well below the EU average.

The latter group of countries clearly lags behind in both the supply and use of data-driven technologies and is not yet in a position to fully exploit the data economy to reap viable business opportunities.

Figure 20 Ranking of Member States by share of Data Users on total Companies, 2016 %
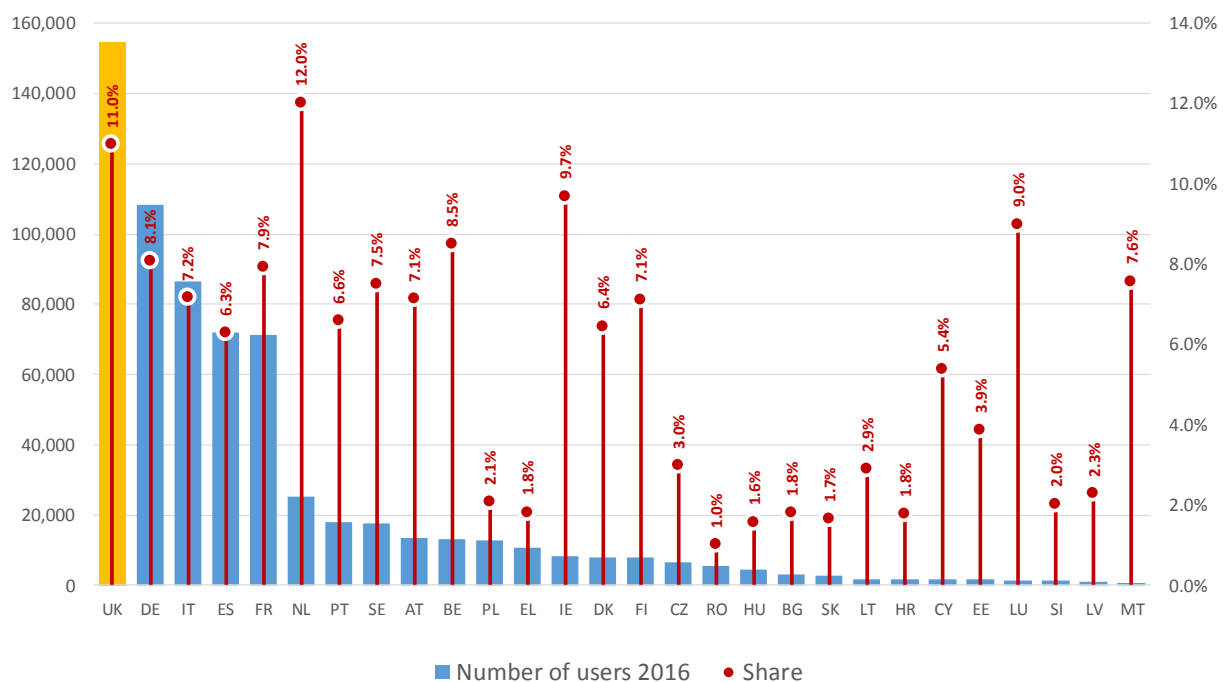
Source: European Data Market Monitoring Tool, IDC 2016

5.4.2 Data Users by Industry

The largest industries by number of data user companies in 2016 were professional services, manufacturing, wholesale and retail and transport totalling together almost two thirds of Europe's population of data users. These industries, however, were not the most dynamic in terms of year-on-year growth as they increased at a growth rate around or slightly below the average EU growth for the years 2016-2015. In contrast, less populated industries such as information and communication technologies, healthcare and finance registered a growth rate well above the EU28 average in 2016, demonstrating their dynamism in terms of adoption and take-up of data-driven technologies.

SMART 2013/0063 European Data Market

Figure 21 Industry Ranking by number of data users, 2015-2016, 000s and growth rates
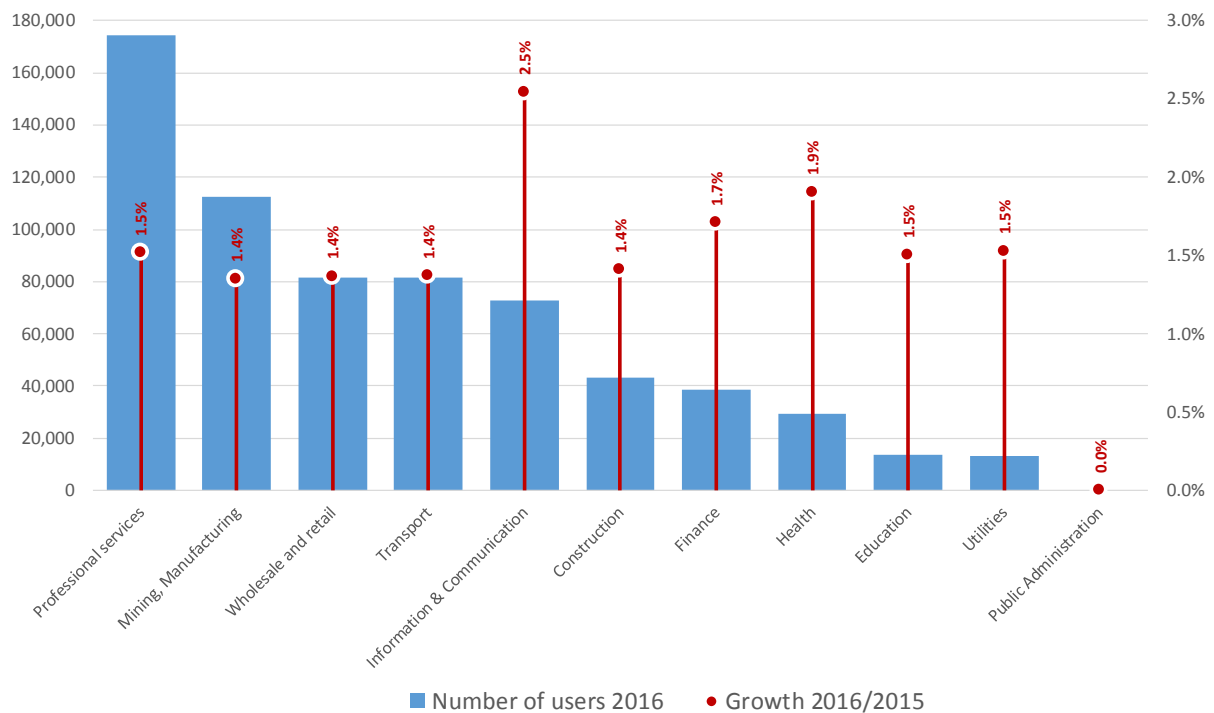
Source: European Data Market Monitoring Tool, IDC October 2016

When looking at data users' shares on total companies by industries, the average EU share of 6.4% in 2016 is largely exceeded in the finance sector, wholesale and retail sector, information and communication technologies and in utilities, that is in industries with a significant intensity of IT use and a large presence of Big-Data and data-based use cases and business opportunities.

On the other side of the spectrum, construction, transport and the public sector confirm their slow pace in adopting data-driven innovation as they present both a relatively low number of data user companies, as well as a very low share of data user companies as a percentage of the total number of companies in their reference industries.

Figure 22 Industry Ranking by share of Data Users on total, 2015, %

■ Number of users 2016   ● Share
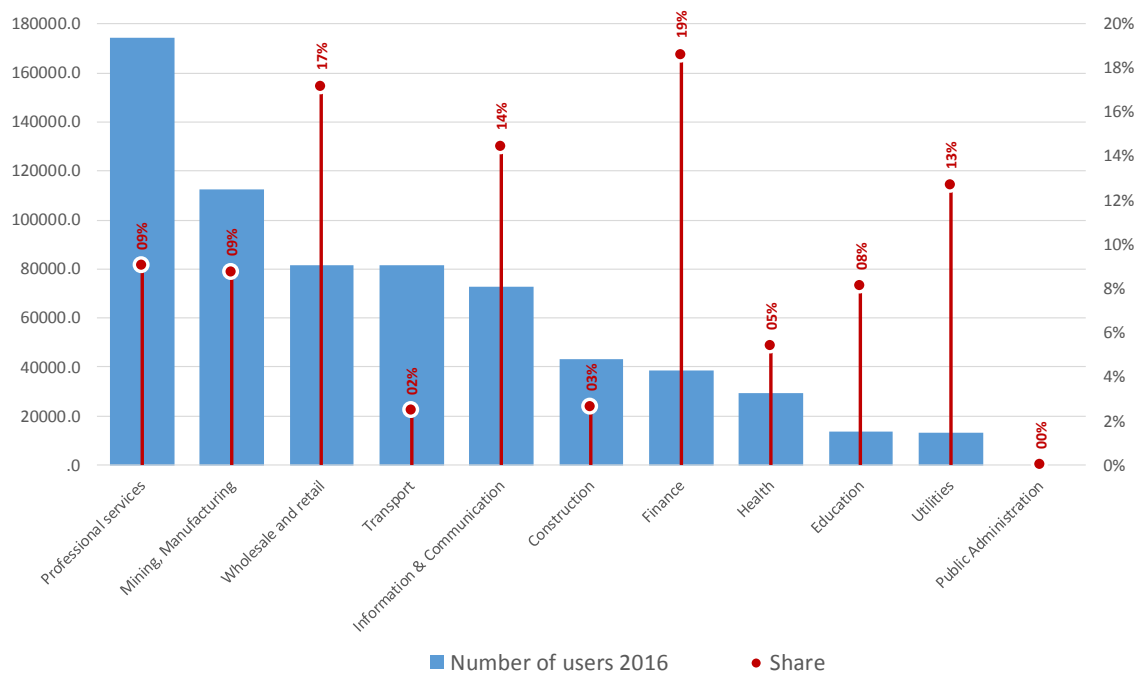
Source: European Data Market Monitoring Tool, IDC October 2016

5.4.3 Data Users by Company Size

In 2016, the SMEs' share of the data user population is very high at 98.9% (in line with the 99% share of SMEs among total enterprises). The penetration of data users stands at 7% for companies with more than 250 employees and at 6.4% for SMEs in 2016, which marks a very moderate improvement from our results in 2015 (up 0.1% in both company sizes year-on-year) and is in line with Europe's industrial structure largely dominated by SMEs.

Table 19 Number and Share of Data Users by Company Size, 2013-2014-2015-2016

	Size Band	1-249 empl.	250+ empl.	Total EU28	SMEs share
Number of data users, Units	2013	627,269	6,336	633,605	99.0%
	2014	636,350	6,650	642,700	99.0%
	2015	643,650	7,100	650,750	98.9%
	2016*	653,800	7,250	661,050	98.9%
Growth rate, %	2015/ 2014	1.1%	6.8%	1.3%	
	2016/ 2015	1.6%	2.1%	1.6%	
Share of total EU companies,	2014	6.3%	13.9%	6.3%	

SMART 2013/0063 European Data Market

	Size Band	1-249 empl.	250+ empl.	Total EU28	SMEs share
%	2015	6.3%	6.9%	6.3%	
	2016*	6.4%	7.0%	6.4%	
Total by company size	2013	10,161,300	46,000	10,207,300	99.5%
	2014	10,172,997	46,053	10,219,050	99.5%
	2015	10,169,650	103,050	10,272,700	99.0%
	2016*	10,223,150	103,550	10,326,700	99.0%

* 2016 estimate

Source: European Data Market Monitoring Tool, IDC October 2016

In line with the overall structure of the EU industry, SMEs critically outnumber large and medium companies. This holds true in the data industry as well, as demonstrated by the SMEs' share of data users, which is slightly lower than the overall share of SMEs in Europe's economy as a whole (98.9% for the former, 99% for the latter in 2016). Medium and large companies do play an important role too, as reflected by the share of data users' companies by size bands, which – according to our latest estimates – amounts to 6.4% for the 1-249 size class and 7% for the 250+ size class. In 2016, the average spending for SME data users is estimated at EUR 27,000 and at more than EUR 6 Million for companies with over 250 employees. The overall average is estimated at EUR 91,000 for data users in 2016, as opposed to the 84,000 estimated for the year 2015. This represent an average growth year-on-year of more than 6% - a remarkable result showing that average spending per user is growing faster than the number of total users, which is typical of an emerging market.

Table 20 Average Data Spending per User Company, by Company Size, 2013-2014-2015-2016, € 000s

Size Band	2013	2014	2015	2016*
1-249	23	24	26	27
250+	5,239	5,498	5,729	6,079
Overall average	75	79	84	91

*2016 estimates

Source: European Data Market Monitoring Tool, IDC October 2016

5.4.4 Data User companies forecast

The number of data user companies is expected to grow steadily over the period 2016-2020 with a CAGR ranging from 0.3% under the Challenge scenario to 13.5% according to the most favourable scenario.

Table 21 Forecast Indicators 2 Data Users, EU 2020, three scenarios

SMART 2013/0063 European Data Market

Indicator 2 .3 and 2.4 – Forecast of Data Users in the EU to 2020, three scenarios							
		Challenge Scenario		Baseline Scenario		High Growth scenario	
N.	Name	Units	CAGR 2016/2020	Units	CAGR 2016/2020	Units	CAGR 2016/2020
2.3	Number of data user companies	668,400	0.3%	727,250	2.4%	1,098,600	13.5%
2.4	Share of data users*	6.5%		7.1%		10.6%	

*: Potential data companies belong to the sectors J – Information and Communication and M – Professional services. The share is based on 2015 numbers sourced from Eurostat

Source: European Data Market Monitoring Tool, IDC October 2016

At Member State level, the concentration of data users will be lower than the concentration of data companies. The UK and Germany will continue to have the lion share of data users under all three scenarios: together with the rest of the largest EU economies (France, Italy and Spain), they will total more than 70% of all data users in 2020 in the Challenge, Baseline and High Growth scenarios. This trend will be particularly pronounced under the latter scenario, which is characterized by a marked take-up of data-related technologies boosting, in turn, the overall population of data users in Europe.

In terms of growth over the period 2016-2020, the picture by Member State is varied and dynamic: medium-sized, advanced and ICT-leading economies such as Sweden, the Netherlands and Belgium, for example, will display above-average growth rates in all the three scenarios under consideration. Larger Member states such as Germany, France and Spain will perform slightly below average in the Baseline scenario but will exhibit on-average growth rates in the Challenge and High-Growth scenarios. Finally, most of the smaller economies (including the Baltic countries, Central and Eastern European Member States and Greece) will grow at a slower pace than the EU28 average across all three scenarios.

Figure 23 Forecast Data Users by Member State by Baseline Scenario, 000s, three scenarios

	Baseline
UK	170400
DE	118450
IT	95700
ES	78700
FR	77850
NL	28650
SE	20050
PT	19800
AT	14750
BE	14500
PL	14100
EL	11600
IE	9150
DK	8700
FI	8600
CZ	7100
RO	5850
HU	4800
BG	3300
SK	2900
LT	2000
EE	1800
HR	1800
CY	1750
LU	1650
SI	1350
LV	1250
MT	700

Challenge Scenario ●——● High Growth Scenario

Baseline Scenario

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market
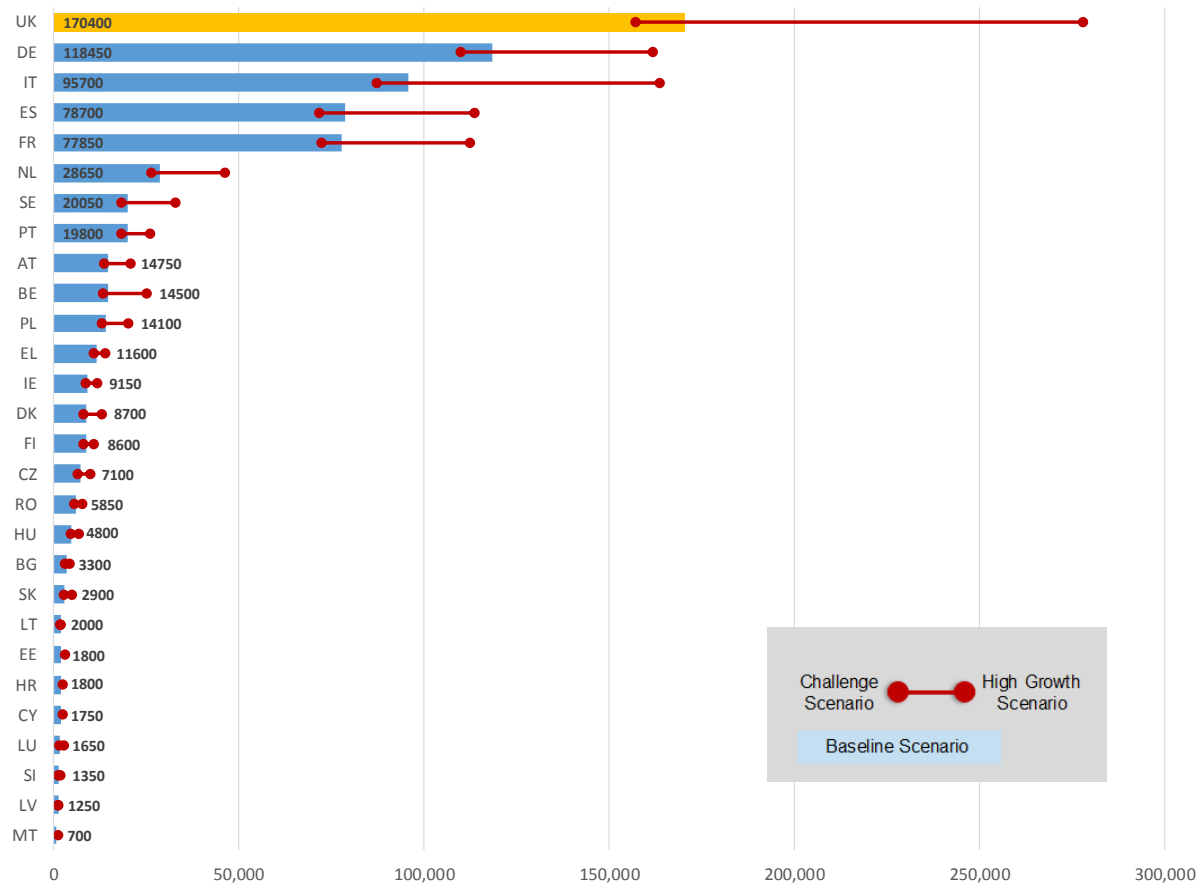
Figure 24 Data Users CAGR 2020/2015 by Member State, ranking by Baseline scenario

SE	3.5%
LV	3.2%
NL	3.1%
SI	3.0%
EE	3.0%
HR	3.0%
BE	2.9%
SK	2.8%
IT	2.5%
HU	2.5%
UK	2.5%
PT	2.4%
LU	2.4%
BG	2.4%
PL	2.3%
DE	2.3%
DK	2.3%
CY	2.3%
ES	2.2%
FR	2.2%
IE	2.2%
EL	2.2%
AT	2.1%
CZ	2.0%
FI	2.0%
LT	2.0%
MT	1.9%
RO	1.8%

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market
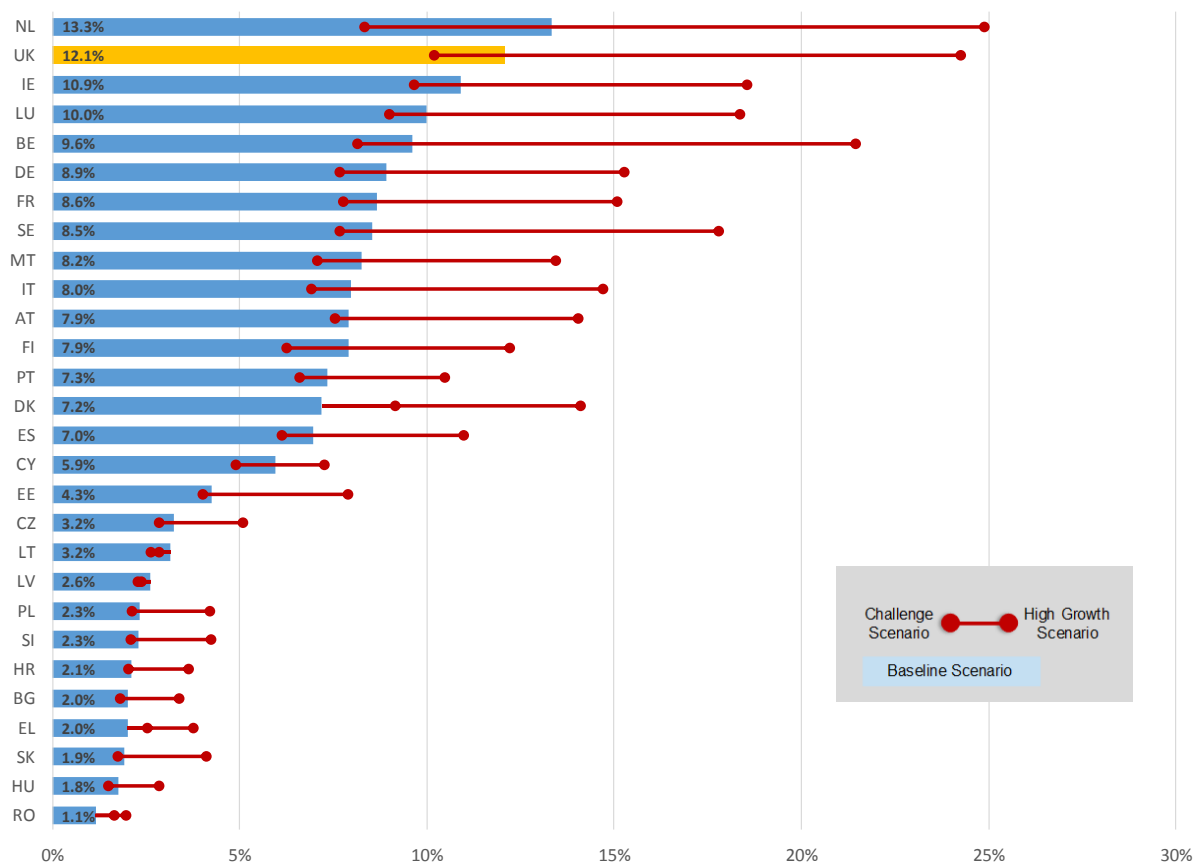
Source: European Data Market Monitoring Tool, IDC October 2016

5.4.5 Data User Companies Forecast by Company Size

SMEs will continue to constitute the vast majority of data users in 2020 under the three scenarios. With a CAGR of 2.4% over the period 2016–2020, the SMEs' share of total data user companies is projected to increase slightly under the High-Growth scenario (99% in 2020) and remain unchanged at 98.9% under the Baseline and Challenge scenarios.

If the data market development is hindered by a slower rate of data-driven innovation, SMEs will grow on a par with the overall 2016–2020 CAGR for all data users and SMEs' share will remain at 98.9%.

Conversely, according to the High Growth SMEs will grow at a CAGR of 13.6% showing a substantial, even more solid in absolute terms, increase of their share of total data user companies.

Table 22 Data User Forecast by Company Size Class, three scenarios

SMART 2013/0063 European Data Market

Number of Data users, units				CAGR of data users, 2020/2015 %		
Size Band	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario	Challenge Scenario	Baseline Scenario	High Growth Scenario
1-249 empl.	661,150	719,400	1,087,700	0.3%	2.4%	13.6%
250+ empl.	7,250	7,850	10,900	0.0%	2.0%	10.7%
Total EU28	668,400	727,250	1,098,600	0.3%	2.4%	13.5%
SMEs Share	98.9%	98.9%	99.0%			

Source: European Data Market Monitoring Tool, IDC October 2016

Table 23 Forecast of average spending by data user by size class

Data Users Average Spending, EUR 000			
Size Band	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
1-249 empl.	33	34	29
250+ empl.	7,120	7,312	6,770
Total EU28	106	111	96

Source: European Data Market Monitoring Tool, IDC October 2016

5.4.6 Key findings

- According to our latest estimates, IDC assesses the population of European enterprises that make strategic and intensive use of datasets and new data technologies at approximately 661,000 in 2016, corresponding to 6.4% penetration of the 10.3 million potential user companies. This is relatively low and consistent with the results for 2013, 2014 and 2015 (6.2% penetration in 2013 and 6.3% in 2014 and 2015) leaving much room to grow in the coming years.
- The UK being one of the strongest ICT markets in Europe with and a leading data-driven economy, it is not surprising to observe a sharp reduction in the numbers of data users in the EU27 vis-à-vis the EU28 across the period under consideration, accompanied by a smoother lowering of the penetration rate which should become of 5.7% in the EU27 in 2016 as opposed to 6.4% in the same year.
- The user population is much less concentrated than the supplier population. The U.K. and Germany have the largest number of users, but together they represent alone nearly 40% of total users. The growth rate from 2015 to 2016 will be of 1.6%, lower than the growth rate for data suppliers, but higher than the growth rate of the period 2014-2015.

SMART 2013/0063 European Data Market

- The largest industries by number of data companies in 2016 were professional services, manufacturing, wholesale and retail and transport, totalling together approximately two thirds of Europe's population of data users.
- In terms of users' shares on total companies by industry, the average EU share of 6.4% in 2016 was largely exceeded in the finance sector, wholesale and retail sector, information and communication technologies and in utilities, that is in industries with a significant intensity of IT use and a large presence of Big-Data and data-based use cases and business opportunities. On the other side of the spectrum, construction, transport and the public sector confirm their slow pace in adopting data-driven innovation as they present both a relatively low number of data user companies, as well as a very low share of data user companies as a percentage of the total number of companies in their reference industries.

6 INDICATOR 3: DATA COMPANIES REVENUES

6.1 Definition

Data companies' revenues correspond to the aggregated value of all the data-related products and services generated by Europe-based companies, including exports outside the EU. This indicator measures the revenues of the data companies identified and classified by Indicator 2, for the products and services specified in our definition of the data market.

The overall value of data companies' revenues is estimated at approximately EUR 62 billion in 2016, with a very healthy growth rate of 10.3% over the previous year; in 2014 the same value was EUR 51.6 billion and slightly exceeded EUR 56 billion in 2015, showing healthy year-on-year growth rates between 8.4% and 10.3%. (Table 24).

The overall value of data revenues is very close, but not identical, to the overall value of the data market for the following reasons:

- The value of the data market corresponds to the aggregated value of all the data-related products and services bought by European users (demand) *including imports from foreign suppliers.*
- The value of revenues corresponds to the aggregated value of all the revenues generated by *Europe-based* enterprises (supply) through the production, distribution, and sale of data-related products and services, *including exports outside of the EU*.

For the sake of simplicity, it is fair to assume that data industry exports and data market imports in the EU are not significant and that tend to balance each other out in a total EU perspective. As in the previous report this does not preclude the existence of a trade balance surplus or deficit at MS level, which we are unable to estimate for lack of specific evidence.

Table 24 Indicator 3 — Total Revenues of Data Companies, EU, 2014–2015-2016

Indicator 3 — Description								
N.	Region	Name	Description	2014 € Mill	2015 € Mill	2016* € Mill	Growth Rate 2015/2014	Growth Rate 2016/2015
3.1	EU27	Total revenues of data companies in the EU	Total revenues of the Data Companies calculated by Indicator 2	40,401	43,609	47,216	7.9%	8.3%
3.1	EU28	Total revenues of data companies in the EU	Total revenues of the Data Companies calculated by Indicator 2	51,686	56,033	61,819	8.4%	10.3%
3.2	EU27	Share of data companies' revenues	Ratio between data companies' revenues and total companies revenues in	3.2%	3.4%	NA	5.8%	7.7%

			sectors J and M					
3.2	EU28	Share of data companies revenues	Ratio between data companies revenues and total companies revenues in sectors J and M	3.2%	3.4%	NA	7.2%	7.4%

* 2016 Estimates

Forecast Indicator 3 – Data Revenues in the EU, 2020								
N.	Region	Indicators	Baseline Scenario		Challenge Scenario		High Growth Scenario	
			2020	16-20 CAGR	2020	16-20 CAGR	2020	16-20 CAGR
3.1	EU27	Total revenues of data companies in the EU	57,633	5.1%	71,453	10.9%	99,347	20.4%
3.1	EU28	Total revenues of data companies in the EU	74,158	4.7%	91,874	10.4%	130,708	20.6%

Source: European Data Market Monitoring Tool, IDC October 2016

6.2.1 Data Revenues by Member State

The size of the data industry is the main element affecting the distribution of data revenues across the EU Member States. As a result, the Member States with the largest economy tend to dominate the picture with the UK, Germany, France, Italy and Spain leading the way ahead of all other European countries.

In terms of growth rates, the UK, Germany, Sweden, Denmark (among others) perform very well (with a growth rate in 2016 well above the 10.3% average at EU28 level) It is worth noting, however, that these are year-on-year variations: as such they reflect only short-term dynamics and have to be considered within the broader framework of long-term growth as displayed by our forecast analysis.

Figure 26 Data Revenues by Member State, 2016, € Million; Growth 2016-2015
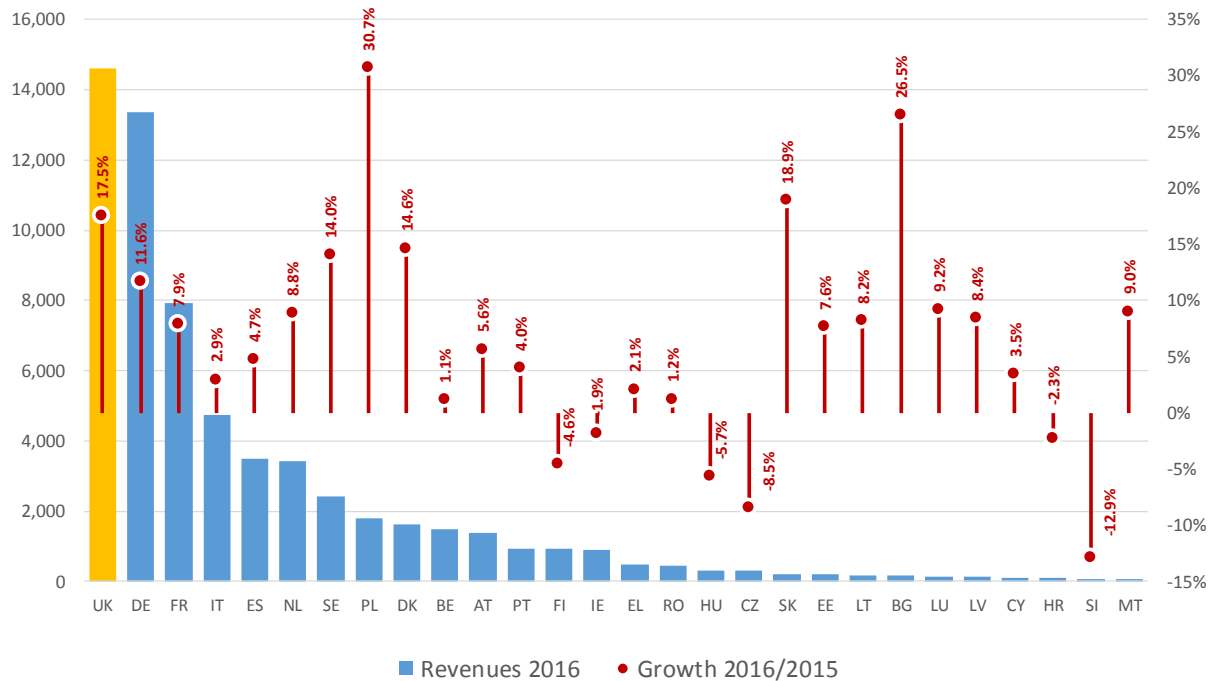
Source: European Data Market Monitoring Tool, IDC October 2016

A key element of comparison is the share of data revenues on the total turnover of the two industries (ICT and professional services) selected as the reference perimeter for the data companies.

The overall turnover of these industries was about EUR 1,626 Billion in 2015. We do not have the same figure for the year 2016, hence we are not in a position to estimate the share of data revenues on the total turnover this year. We estimated the total data revenues for the EU28 to represent 3.4% of the industries' turnover of information and communication and professional services (J and M of the NACE rev2 codes of the economic activities) – up 0.2% form our previous estimates for the year 2014.

SMART 2013/0063 European Data Market
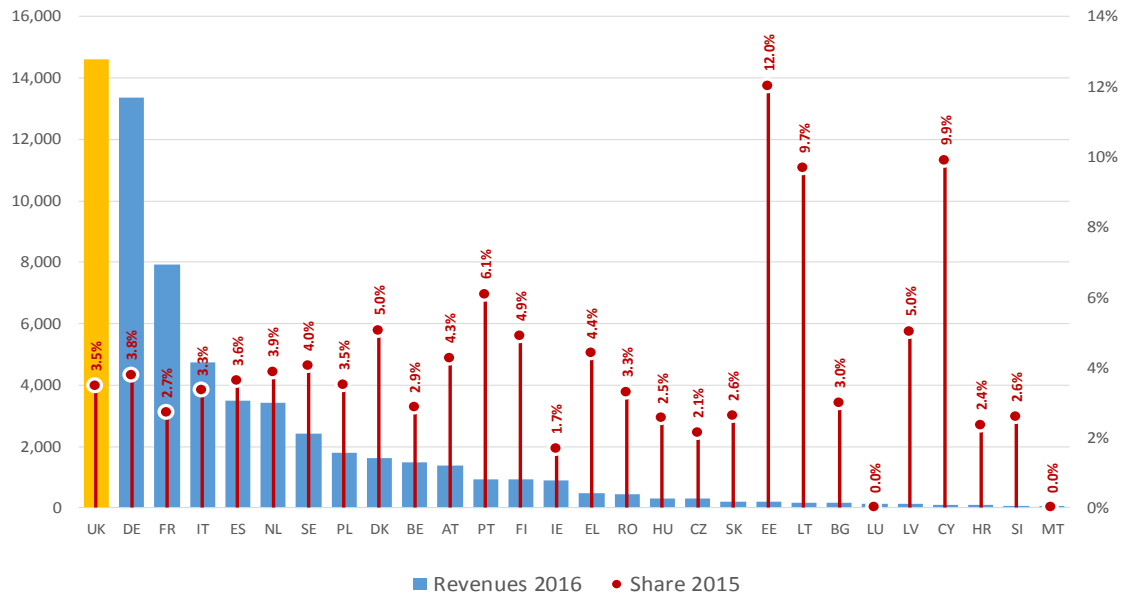
% Base: Turnover EU28 NACE J and M 70 72 73 74 for companies with 0+ empl. by MS

Source: European Data Market Monitoring Tool, IDC October 2016

6.2.2 Data Revenues by Company Size

The estimate of data revenues by company size is articulated in four size classes instead of two in order to provide a more accurate picture of the differences between companies, as very small enterprises behave differently from large enterprises but they are also very numerous.

Based on our analysis, small enterprises between 1 and 49 employees will capture approximately 55% of revenues in 2016, which is near to the 2015 estimate and shows a yearly growth rate of 10.5% across the period. Their average revenues are also on the increase and are estimated at EUR 149,000 in 2016, EUR 137,000 in 2015, 128,000 in 2014, and EUR 120,000 in 2013. This group of companies tends to include small software and IT services companies, but also startups and very young SMEs. Their relative share of revenues is smaller than their share of the total population of enterprises, but this is normal since the data industry is in its early stages of development and for some Big Data technologies large IT players moved quickly into the market.

Medium-sized enterprises with up to 249 employees represented 18% of total data revenues in 2015, with an average revenue of EUR 430,000 per year and a growth rate of 8.6%. In 2016, their share of total data revenues is substantially unvaried, with an average revenue of EUR 438,000 in 2016. In 2014, the same revenue was EUR 400,000 and in 2013 EUR 300,000 – a constant increase over the four years under consideration. This group of companies is very dynamic and includes some of the most innovative SMEs in the data market.

In 2016 the 250-499 size class follows a similar dynamic, showing 25% of revenues and average revenues of EUR 5.6 Million per year, and 11% growth from 2015 (EUR 5.2 Million in 2015, EUR 5 Million in 2014 and EUR 4.7 Million in 2013).

Finally, the largest size class of companies represents 2.4% of total revenues; in absolute terms, each company in this size class will generate more than EUR 17 Million of sales on average in 2016 vis-à-vis EUR 16 Million in 2015, EUR 15.1 Million in 2014 and EUR 14 Million in 2013.

Based on our analysis the average data revenue per company was EUR 219,000 in 2013, EUR 212,000 in 2014, increasing to EUR 225,000 in 2015 and amounting to EUR 243,000 in 2016. The cumulative revenue share of SMEs remains at around 72% across the four years under consideration. We expect that smaller enterprises will catch up with the larger ones in the coming years as the data industry continues to mature.

Table 25 Total Data Companies Revenues by size class, 2013-2014–2015-2016

	Size Band	1-49 empl.	50-249 empl.	250-499 empl	500+ empl.	Grand Total	SMEs share
Total data revenues € Mill	2013	25,755	8,418	12,381	1,173	47,801	71.6%
	2014	27,796	9,200	13,375	1,285	51,686	71.6%
	2015	30,629	9,989	14,049	1,362	56,033	72.5%
	2016*	33,842	10,915	15,569	1,494	61,819	72.4%
Total data revenues % on total	2014	54%	18%	26%	2%		
	2015	59%	19%	27%	3%		
	2016*	65%	21%	30%	3%		
Growth rate total revenues	2015/2014	10.2%	8.6%	5.0%	6.0%	8.4%	
	2016/2015	10.5%	9.3%	10.8%	9.7%	10.3%	
Share of data revenues on total revenues	2013	4.8%		1.5%		3.0%	
	2014	5.1%		1.6%		3.2%	
	2015	5.4%		1.7%		3.4%	
	2016*	NA		NA		NA	

*2016 Estimates

Source: European Data Market Monitoring Tool, IDC October 2016

Table 26 Average Data Companies Revenues by size class, 2013-2014–2015-2016

Size Band	Average revenues data per company, 2014 € 000	Average data revenues per company, 2015 € 000	Average data revenues per company, 2016 € 000
1-49 empl.	128	137	149
50-249 empl.	400	430	438
250-499 empl.	5,000	5,200	5,634
500+ empl.	15,100	16,000	17,238
Grand Total	212	225	243

Source: European Data Market Monitoring Tool, IDC October 2016

6.2.3 Key Findings

- In the EU28 the overall value of data companies' revenues is estimated at approximately EUR 62 Billion in 2016, with a healthy growth rate of 10.3% over the previous year. Data companies' revenues were more than EUR 56 Billion in 2015, EUR 51 Billion in 2014 and EUR 47.7 Billion in 2013 - a constant positive trend throughout the period.
- Based on our analysis the average data revenues per company were EUR 199,000 in 2013, EUR 212,000 in 2014, increasing to EUR 225,000 in 2015 and raising to 243,000 in 2016.
- The cumulated revenue share of SMEs is at around 71-72%%, which is stable throughout the period 2013-2016.
- Average revenues are growing for all size classes, with the smallest companies growing slightly slower than the others. Our expectation is that the smaller enterprises will catch up with larger ones in the coming years as the data industry continues to mature.
- Medium-sized enterprises with 50 to 249 employees represent more than 10% of data companies, but captured almost 21% of total data revenues in 2015 (up 2% points from the previous year), with an average revenue of EUR 438,000 per year and growth of 9.3%. They are the companies to watch in this market.

6.2.4 Data Revenues forecast by Member State

Data companies' revenues will experience healthy growth over the next few years according to all three scenarios under consideration. According to our Baseline scenario, in the EU28 revenues will increase of 10.4% from 2016 to 2020, whereas the pace will reduce to 4.7% under the Challenge scenario and reach a considerable 20.6% if the assumptions underlying the High Growth scenario materialize.

The data revenues' forecast at 2020 without the UK (EU27) will not undergo significant changes across all three scenarios under consideration. In the EU27 only the High-Growth scenario would be negatively affected by Brexit registering a CAGR of 20.4% over the period 2016-2020 instead of 20.6% in EU28.

In terms of Member State dynamics, the largest EU economies dominate the scene under all three scenarios with the UK, Germany, France, and Italy coming at the top of the league by revenues amount in 2020 in the Baseline, Challenge and High Growth scenarios. Instead, when looking at the CAGR 2016-2020, smaller economies take the lead with countries like Belgium, Sweden, the Netherlands, but also Croatia, Slovenia, the Czech Republic Slovakia displaying growth rates well above the EU average under all three scenarios.

Figure 28 Data revenues forecast by Member State, 2020, ranking by High Growth scenario, EUR Million

Source: European Data Market Monitoring Tool, IDC, October 2015

SMART 2013/0063 European Data Market
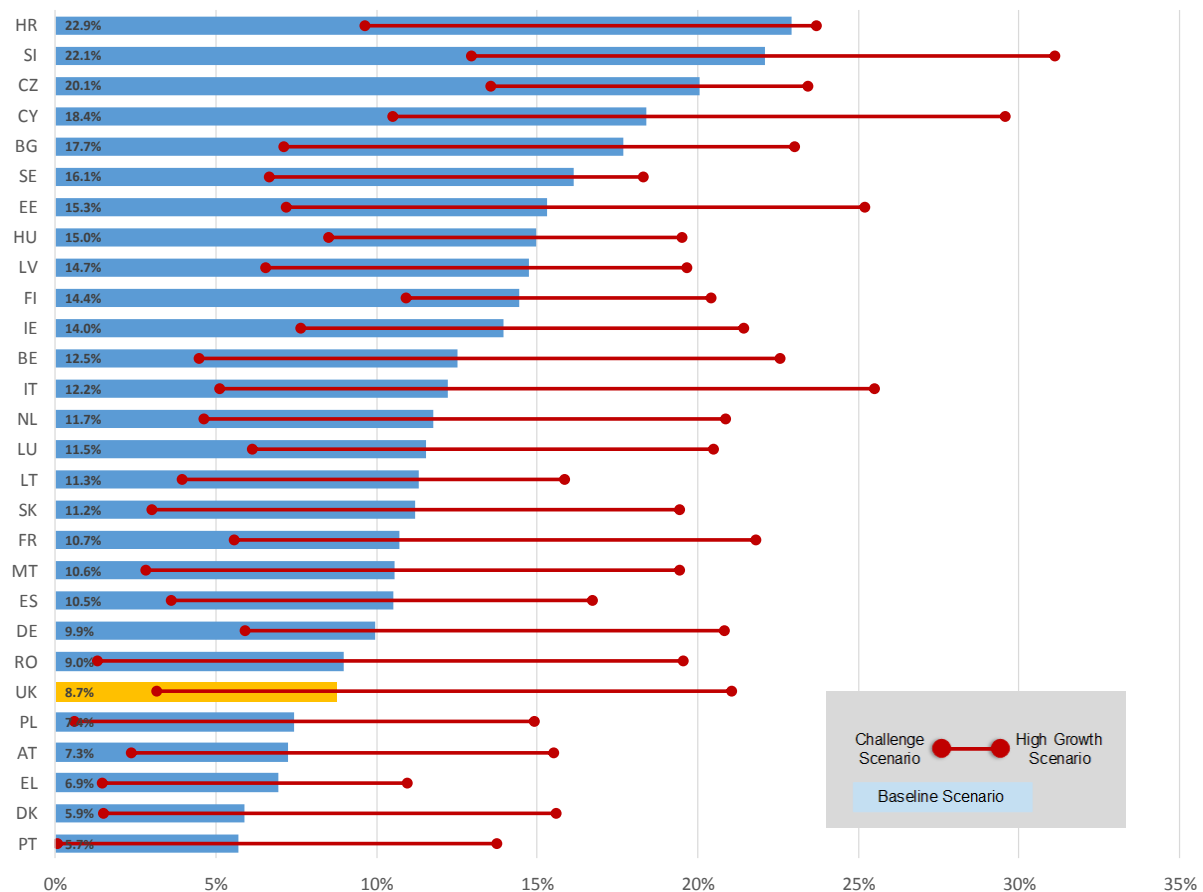
Source: European Data Market Monitoring Tool, IDC October 2016

6.2.5 Data Revenues Forecast by Company Size

The revenues of data companies will increase in the next four years in all the three scenarios under consideration. Table 27 below summarizes the data revenues forecast by company sizes bands under the Challenge, Baseline and High Growth scenarios.

Table 27 Data revenues forecast by company size band, 2016; 2016-2020, € Million

Data Companies Revenues EUR Million				
Size Band	2016	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
1-49 empl.	33,842	39,043	47,730	64,064
50-249 empl.	10,915	13,241	17,142	26,967
250-499 empl.	15,569	19,740	24,347	36,292

SMART 2013/0063 European Data Market

Data Companies Revenues EUR Million				
Size Band	2016	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
500+ empl.	1,494	2,134	2,654	3,385
Total EU28	61,819	74,158	91,874	130,708
SMEs share	72.4%	70.5%	70.6%	69.6%

Source: European Data Market Monitoring Tool, IDC, October 2016

Table 28 Forecast of average data revenues per company per company size, 2016; 2016 - 2020, € thousand

Data Companies Average Revenues EUR thousand				
Size Band	2016	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
1-49 empl.	149	165	172	199
50-249 empl.	438	511	576	795
250-499 empl.	5,634	6,863	7,237	9,204
500+ empl.	17,238	23,656	24,890	26,986
Total EU28	243	280	296	364

Source: European Data Market Monitoring Tool, IDC October 2016

6.2.6 Key Findings

- The revenues of the data companies by Member State will expand significantly over the next four years in all three scenarios, although at a slower pace in the Challenge scenario (4.7% CAGR versus 10.4% in the Baseline scenario and a buoyant 20.6% under the High Growth scenario.
- The largest EU economies will dominate the scene in the year 2020 in terms of the absolute amount of data revenues with the UK, Germany, France and Italy prevailing and coming at the top of the list in all three scenarios. Nevertheless, smaller economies (e.g. Belgium, Sweden and the Netherlands) will take the lead in terms of relative growth rates in the period 2016-2020.
- Data revenues grow faster than the market and the number of data companies, because we expect them to increase their average revenues in time, after they enter the industry.

SMART 2013/0063 European Data Market

7. MEASURING THE DATA MARKET

7.1 Definition

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

In the present study, the data market captures 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 (please see indicator 4.2 "Value of the Data Economy"). Further, the data market represents a wider concept than the market of Big Data & Analytics (BDA) as it includes not only the value generated by pure data players developing BDA technologies but also the value created by data-related research, businesses, information and IT services. The digital data exchanged as "products" or "services" in the data market refer exclusively to data that 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, image databases). The value of the data market is not exactly equal to the aggregated revenues of the European data companies because it 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 29 provides an overview of the Data Market value in Europe over the past four years and its forecast at 2020 according to the three scenarios under consideration.

Table 29 Indicator 4.1 Value and Growth of the Data Market

Indicator 4.1 — Value and Growth of the Data Market (€ Million; %)								
N.	Market	Name	Description	2013	2014	2015	2016	Growth rate 2016/2015
4.1	EU27	Value of the Data Market	Estimate of the overall value of the Data Market	38,022	41,012	42,586	46,226	8.5%
4.1	EU28	Value of the Data Market	Estimate of the overall value of the Data Market	47,419	50,888	54,351	59,539	9.5%

SMART 2013/0063 European Data Market

							CAGR 2020/ 2016 Challenge scenario	CAGR 2020/ 2016 Baseline scenario	CAGR 2020/ 2016 High Growth scenario
N.	Market	Name	Description	2020 Challenge (000)	2020 Baseline (000)	2020 High Growth (000)			
4.1	EU27	Value of the Data Market	Estimate of the overall value of the Data Market	54,390	61,889	80,632	4.1%	7.6%	14.9%
4.1	EU28	Value of the Data Market	Estimate of the overall value of the Data Market	70,407	79,637	106,821	4.3%	7.5%	15.7%

Indicator 4.1 — Value and Growth of the Data Market (€ Million; %)

Source: European Data Market Monitoring Tool, IDC October 2016

The European data market in the EU28 is now estimated at EUR 54,351 million in 2015 and at EUR 59,539 million in 2016, thus exhibiting a solid year-on-year growth of 9.5%. This positive trend will continue throughout the next four years, although at different paces according to the selected scenarios, registering a 7.5% growth rate under the Baseline scenario, a 15.7% under the High Growth scenario and a 4.3% under the Challenge scenario. If we exclude the UK from our projected estimates, the value of the data market will grow at the same pace in the EU27, although slightly slower in both the Challenge scenario (with a CAGR of 4.1% vs. a CAGR of 4.3% in the EU28) and in the High-Growth scenario (with a CAGR of 14.9% vs. 15.7% in the EU28). As one of the most vibrant ICT-driven economies, the UK confirms its importance in an economy which is more and more driven by data and data-related products and services. As a result, its exclusion will have a negative (although minor) impact on the overall dynamics of the European data market.

As a percentage of the total ICT spending in the EU28, the data market is now estimated to represent a share of 9.5% - against a share of 8.8% in 2015 – and is expected to improve under all of the three scenarios considered in this study with respect to our previous estimates (14.8% under the Baseline scenario; 11.1% under the Challenge scenario and 14.5% in the High Growth scenario). The reason for this positive development is explained by an overall ICT spending as a whole that is likely to grow at a more moderate pace in the years to come, while the data-related components of its spending (i.e. the data-driven technologies such as Big Data and analytics) will increase at a faster pace. Overall the data market is likely to go through slight variations following the UK vote to leave the European Union in June 2016. The data market is an emerging market and, as such, it will continue to show high early growth. In addition, any changes to demand in the UK will most likely be taken up by other member states as external investors are likely to switch member states for their location rather than abandon the EU as a market. The three forecasts for 2020 – Baseline, Challenge, and High-Growth - reflect partially the impact of the UK vote.

The Data Market and the Member States' Dynamics in 2016

The data market value distribution by Member State in 2016 is exhibited in Figure 31. Five Member States alone (Germany, the UK, France, Italy and the Netherlands) muster 70% of the total data market value in the EU28 in 2016.

Figure 30 Data Market Value (€M) and Share of Data Market Value (%) by MS, 2016
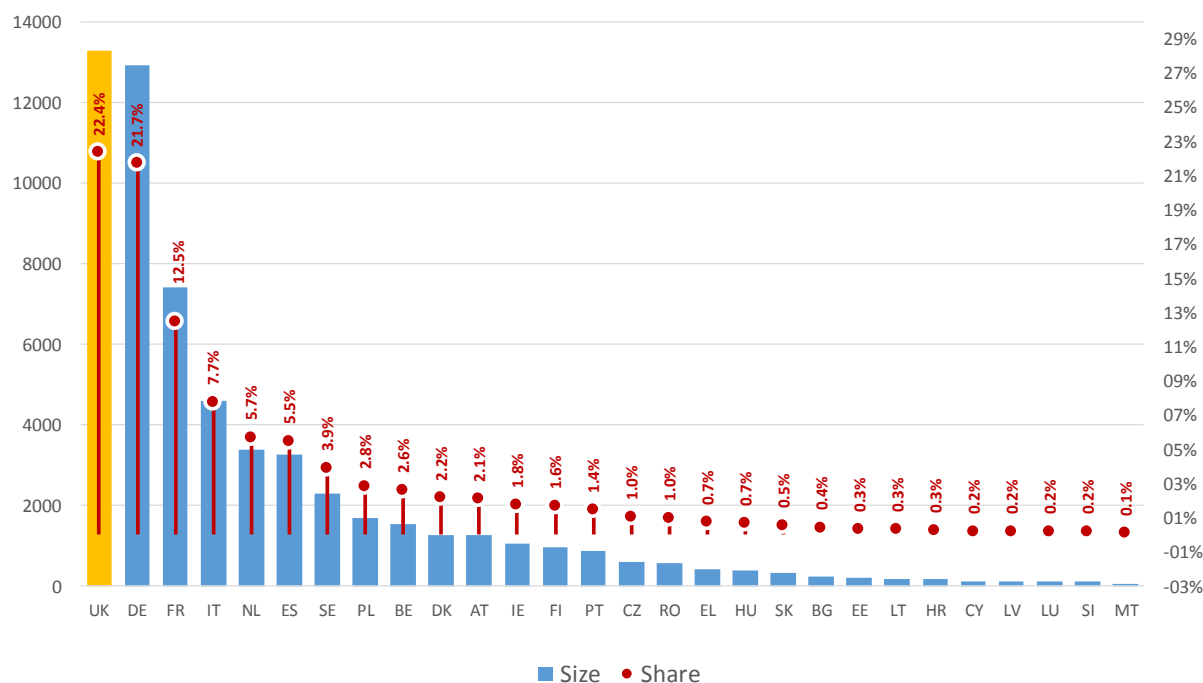
Source: European Data Market Monitoring Tool, IDC October 2016

Most of the leading EU economies will continue to exhibit solid growth in 2016 with respect to 2015 – this is notably the case of Germany (growing 10% year-on-year), France (7.5% year-on-year) and Spain (4.1% year-on-year) but also of smaller and dynamic ICT-oriented economies such as Ireland (growing 18.8% year-on-year) and Denmark (14.2% year-on-year). Interestingly, less mature economies such as Croatia, the Czech Republic, Hungary, Romania, Slovakia and the Baltic countries will all witness a healthy growth in their data market dynamics with growth rates between 11% and 17% year-on-year. While some of these Member States were already amongst the most dynamic players in the data market scene (it is the case of the Baltic countries, for example), others, especially the Member States from Central and Eastern Europe, are confirming the "catching-up" trend that was evident in our previous versions of the present study.
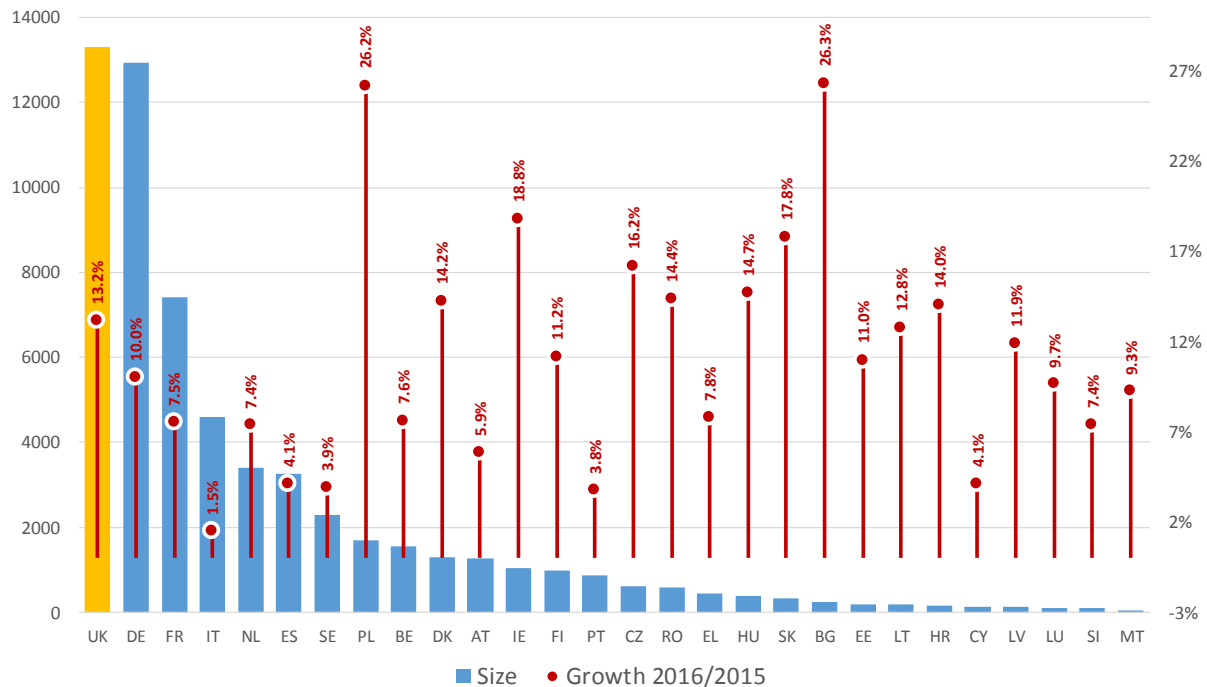
Source: European Data Market Monitoring Tool, IDC October 2016

Another way to look at these variables and further investigate the data market dynamic at Member State level is to compare the relative share of the data market on the total EU28 with the relative ICT spending share as a percentage of the total EU ICT spending. As shown in Figure 31 below, some Member States present a relative data market share that is still lower than their ICT spending share. In other words, the country's relative weight at EU level in terms of data market is lesser than its relative weight in terms of ICT spending. In these Member States, the data market is expected to grow at fast pace in order to catch up with the ICT developments. This is notably the case of some of the largest EU economies such as France, Italy and Spain but also of smaller and dynamic economies such as the Czech Republic and Hungary.

On the other hand, where the relative market share is higher than the relative ICT spending share, the data market has already achieved a significant level of development and it is likely to evolve steadily, although possibly at a slower speed, than in other Member States. As in our previous analyses, this is the case of Germany and also of other countries such as Ireland, the Netherlands, Sweden and, to a lesser extent, Austria and Belgium.

Figure 32 Data Market Share by MS (%) and ICT Spending Share by MS, 2016

Legend: ■ DM Share 2016   ● ICT Share 2016
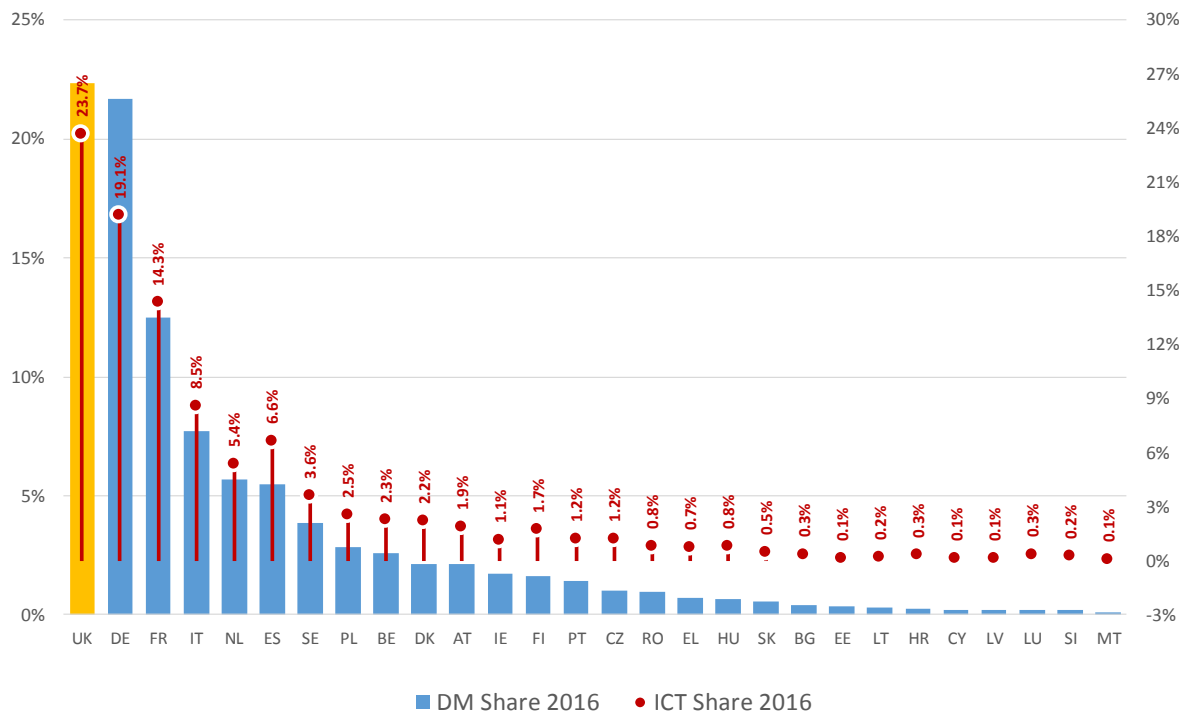
Source: European Data Market Monitoring Tool, IDC October 2016

According to IDC's most recent research, the overall ICT spending in the EU in 2016 has suffered a halt and is now expected to grow at a much slower pace than what was estimated last year. The double-digit growth experienced over the period 2014-2015 will turn to a much more modest grow of 0.9% in 2016-2015 and the same applies for 2020 projections where the overall ICT spending's growth for the period 2016-2020 is expected to exhibit a Compound Average Growth Rate (CAGR) of approximately 1 percentage point lower than our previous estimates (0.4% vs. 1.8% for the Baseline scenario; -3.7% vs -2.9% for the Challenge scenario; 4.3% vs 5.2% for the High-Growth scenario). In contrast, the value of the data market shows a more positive trend over the same period. As a result, the data market shares on ICT spending in 2016 has been revised upwards and is now at 9.5% for the EU28 (up of almost 2 percentage points with respect to our previous forecast) and at 9.7% for the EU27 if we exclude the UK.

The figure below summarizes the most recent data market shares on ICT spending by Member State.

Figure 33 Data Market Share on ICT Spending by Member State, 2016, %
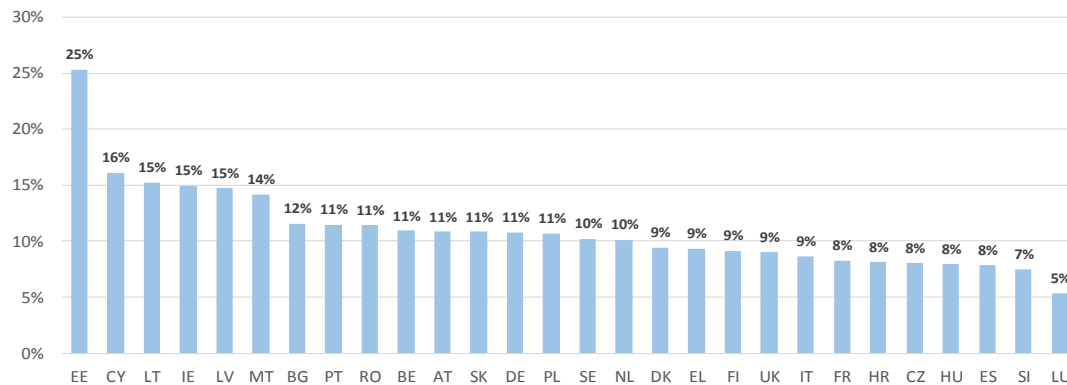
Source: European Data Market Monitoring Tool, IDC October 2016

The Data Market and the Industry Dynamics in 2016

Figure 35 displays the value of the data market by industry and the relative growth by industry year-on-year across the EU28. The top industries in terms of data market size continue to be represented by industry sectors that make a significant usage of data-related technologies, i.e.: the manufacturing sector, the finance industry, the area of professional services and the ICT sector. Not surprisingly, the latter is also one of the sectors exhibiting one of the highest growth rates in 2016 with respect to the previous year. Apart from the ICT sector, as in our previous analysis, the most dynamic growth by industry in 2016 is performed by vertical markets that still hold a relatively small size of the overall data market spending by industry. This proves that the data-related technologies are rapidly finding new ground in previously unchartered areas and are growing fast in sectors like home, healthcare activities, education and the public sector as whole.

Figure 34 Data Market Value by Industry, 2016, € Million; Data Market Growth by Industry, 2016-2015, %

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market
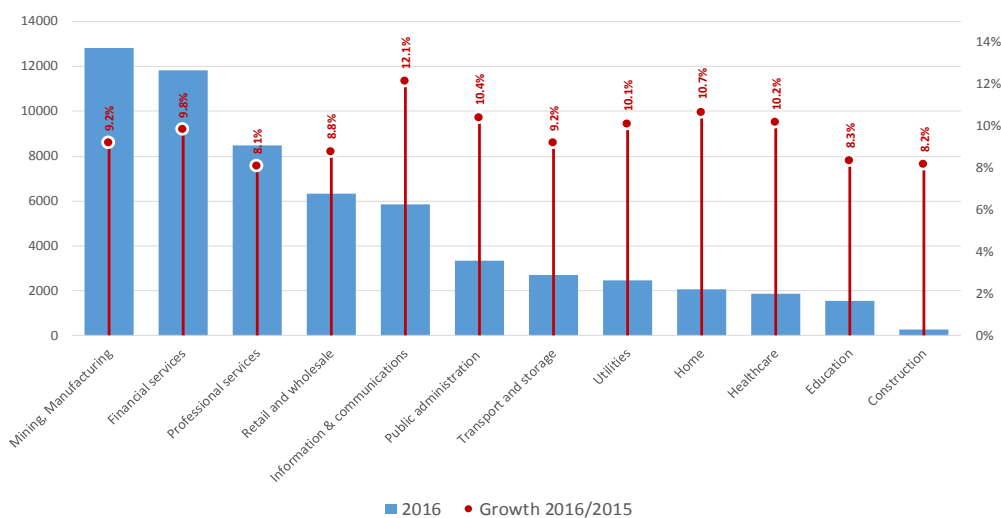
The high concentration rate of the data market by industry is confirmed by our latest analysis: Manufacturing, Financial Services and Professional Services alone make up more than 50% of the overall data market value in 2016 (see Figure 34, below).

Figure 35 Data Market Value (€M) by Industry, 2016; Data Market Share (%) by Industry, 2016
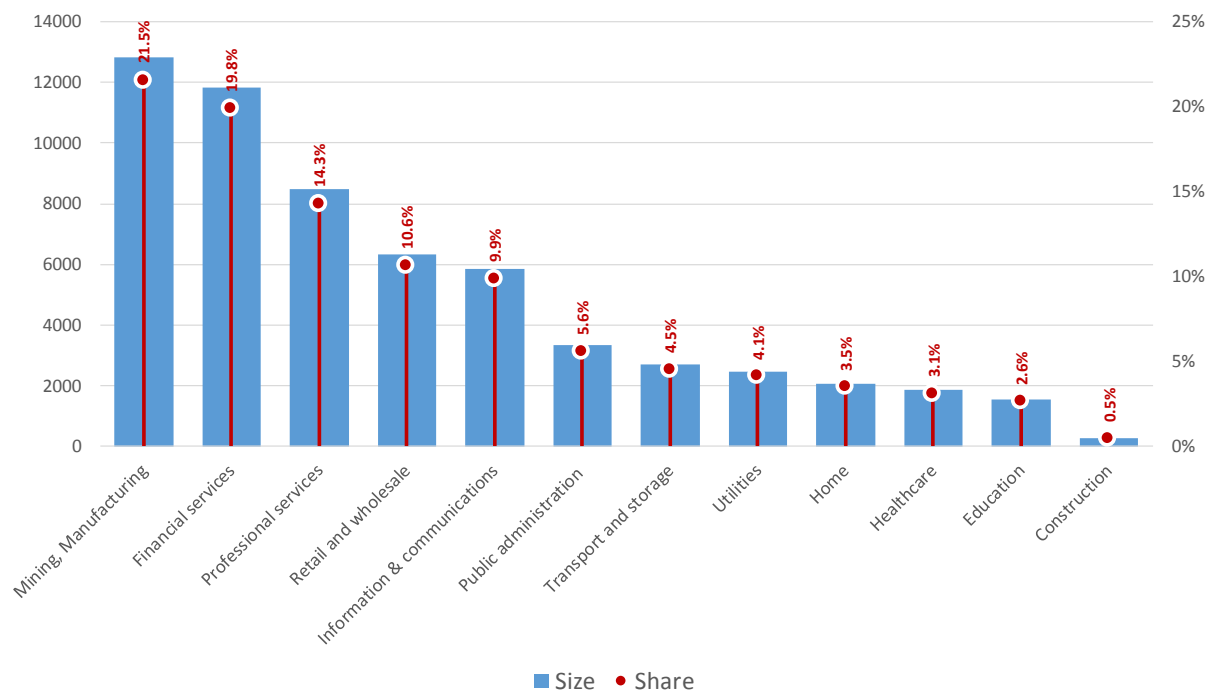
Source: European Data Market Monitoring Tool, IDC October 2016

A somehow more interesting picture emerges when comparing the relative share of each industry in terms of the total data market spending with their respective relative share in terms of total ICT spending. The results for the year 2016 across the 12 industries under consideration are portrayed in Figure 35 below.

Similar to what we experienced by Member State, the most interesting cases of data market dynamics by industry are where the relative share of data market spending is lower than the relative share of ICT spending as this is where potential growth is most likely to happen in the near future.

Conversely, where the relative share of data market spending is already higher than the relative share of ICT spending, data market growth tends to rely more on non-ICT factors, thus theoretically slowing down the growth pace of the data market as a whole.

Figure 36 Data Market Share by Industry (%) and ICT Spending Share by Industry, 2016

SMART 2013/0063 European Data Market
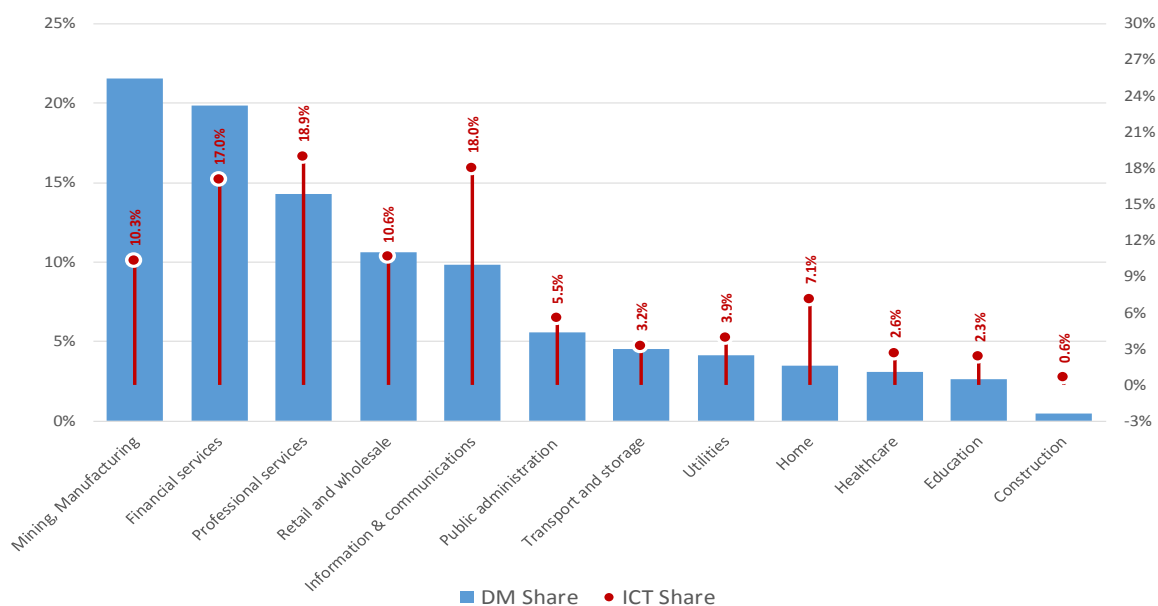
Source: European Data Market Monitoring Tool, IDC October 2016

To further investigate the data market dynamics by industry, we present below our latest calculation of the share of the data market on the total of ICT spending in EU28 by industry. In 2016 the overall spending for the data market in the EU28 represented 9.5% of the total ICT spending in the EU28.

Figure 37 Data Market Share on ICT Spending by Industry, 2016, %
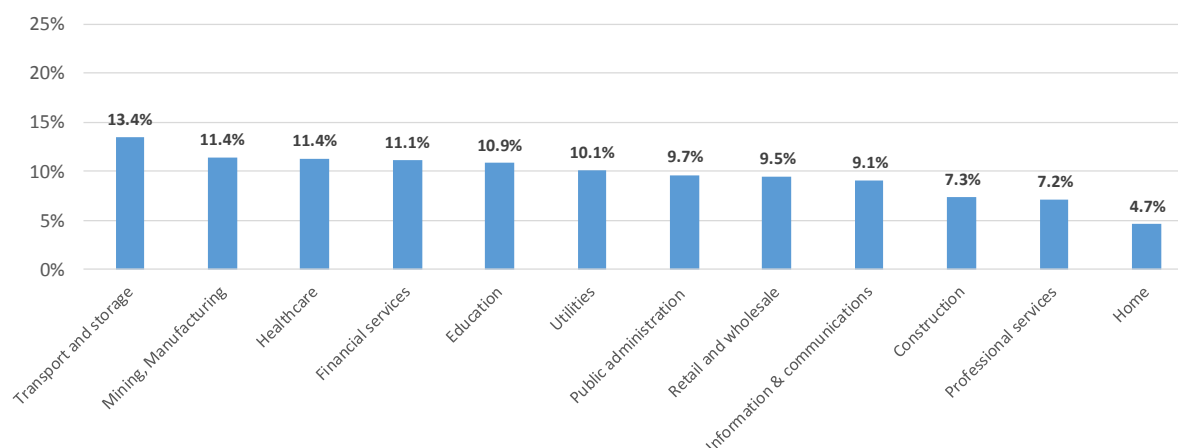
Source: European Data Market Monitoring Tool, IDC Octobe2016

Figure 38 shows a slightly changed picture from what we produced in the course of our previous analyses. As usual, the industries contributing the most to the overall ICT spending in Europe are often those displaying a higher data market spending with respect to the EU28 average in the EU. However, given the recent slowdown of the ICT spending dynamic in Europe, investments in data-related technologies (and the subsequent production of data-related products and services) has gone through a partial

117

SMART 2013/0063 European Data Market

reshuffle to the advantage of traditional industries such as Transport Mining and Manufacturing, and Healthcare (all with a share of data market well above the EU28 average) and to the detriment of ICT-oriented sectors such as Information & Communications and Professional services, which now exhibit a data market share on ICT slightly below the EU28 average.

7.2 Indicator 4.1: Data Market Forecasts

The value of Europe's data market in 2020 will continue to grow steadily according to our latest projections. Nevertheless, due to the relative slowdown of the overall ICT spending and investments expected over the next four years, our estimates by scenarios have been slightly revised and now display a slightly weaker growth in the Challenge and Baseline scenarios (4.3% vs. 5.8% and 7.5% vs 9% respectively). The High-Growth scenario, conversely, displays a similar growth rate than our previous estimate (15.7% vs. 15.5%), showing that the possible negative effects of Brexit on the overall data market value will be recouped by 2020 and that the overall investments in data-related technologies will resume solid growth if the assumptions underlying this positive scenario prove true.

At Member State level our forecast for the period 2015-2020 under all three scenarios does not show any significant change to the picture we have drawn for 2015–2016. Figure 37 portrays the absolute size of the data market by Member State under the three scenarios. As outlined in the previous report, the size of the data market will continue to be closely interconnected with the Member States' relative economic strength and the size of their ICT market. As a result, the U.K., Germany, France and Italy will continue to hold the lion's share of the data market under the three scenarios.

Figure 38 Data Market Value by Member State, 2020, € Million, three scenarios

UK	17748
DE	16359
FR	9134
IT	6327
NL	5609
SE	4450
ES	4045
BE	2340
PL	2148
DK	1637
AT	1508
IE	1270
FI	1154
PT	1118
CZ	722
RO	636
HU	558
EL	540
SK	488
BG	344
EE	285
LT	249
HR	227
CY	180
SI	169
LV	161
LU	153
MT	78

Challenge Scenario ●———● High Growth Scenario

Baseline Scenario
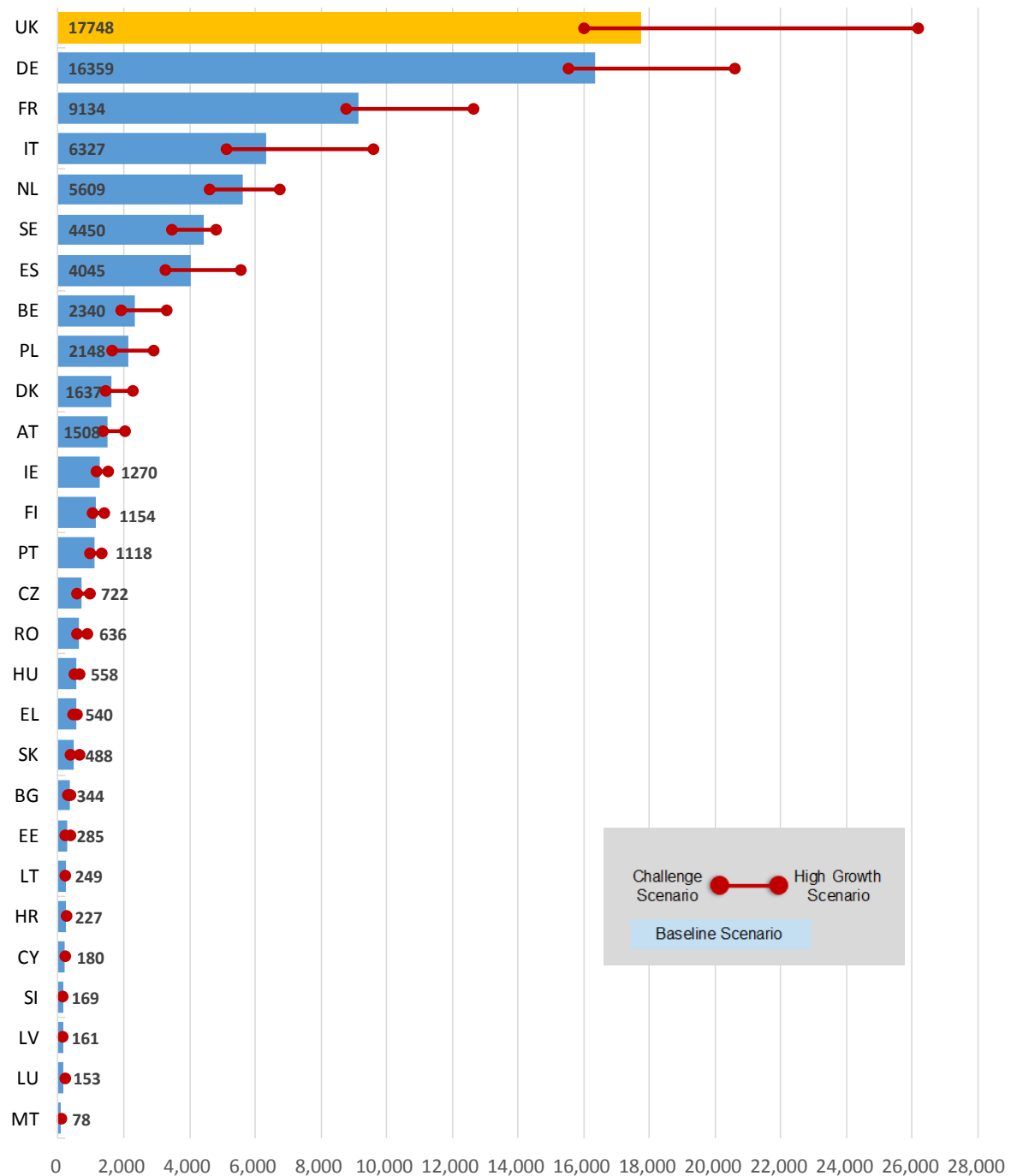
Source: European Data Market Monitoring Tool, IDC October 2016

A closer look at the Member States' growth dynamic reveals that the countries with a relatively high percentage of data market over their total ICT spending in the period 2015–2016 (see Figure 40) will continue to perform very well during the years 2017 to 2020.

119

Figure 39 Data Market Growth 2016-2020 by Member State, 2020, %, three scenarios

Member State	Baseline
SE	17.9%
NL	13.4%
BE	11.0%
SK	10.9%
CY	10.7%
SI	10.6%
MT	10.0%
EE	9.6%
HU	8.9%
HR	8.5%
BG	8.5%
LT	8.3%
IT	8.3%
LV	8.1%
LU	7.7%
UK	7.5%
PT	6.8%
DK	6.3%
PL	6.2%
DE	6.1%
EL	5.9%
ES	5.5%
FR	5.3%
IE	5.1%
AT	4.6%
CZ	4.2%
FI	4.2%
RO	2.2%
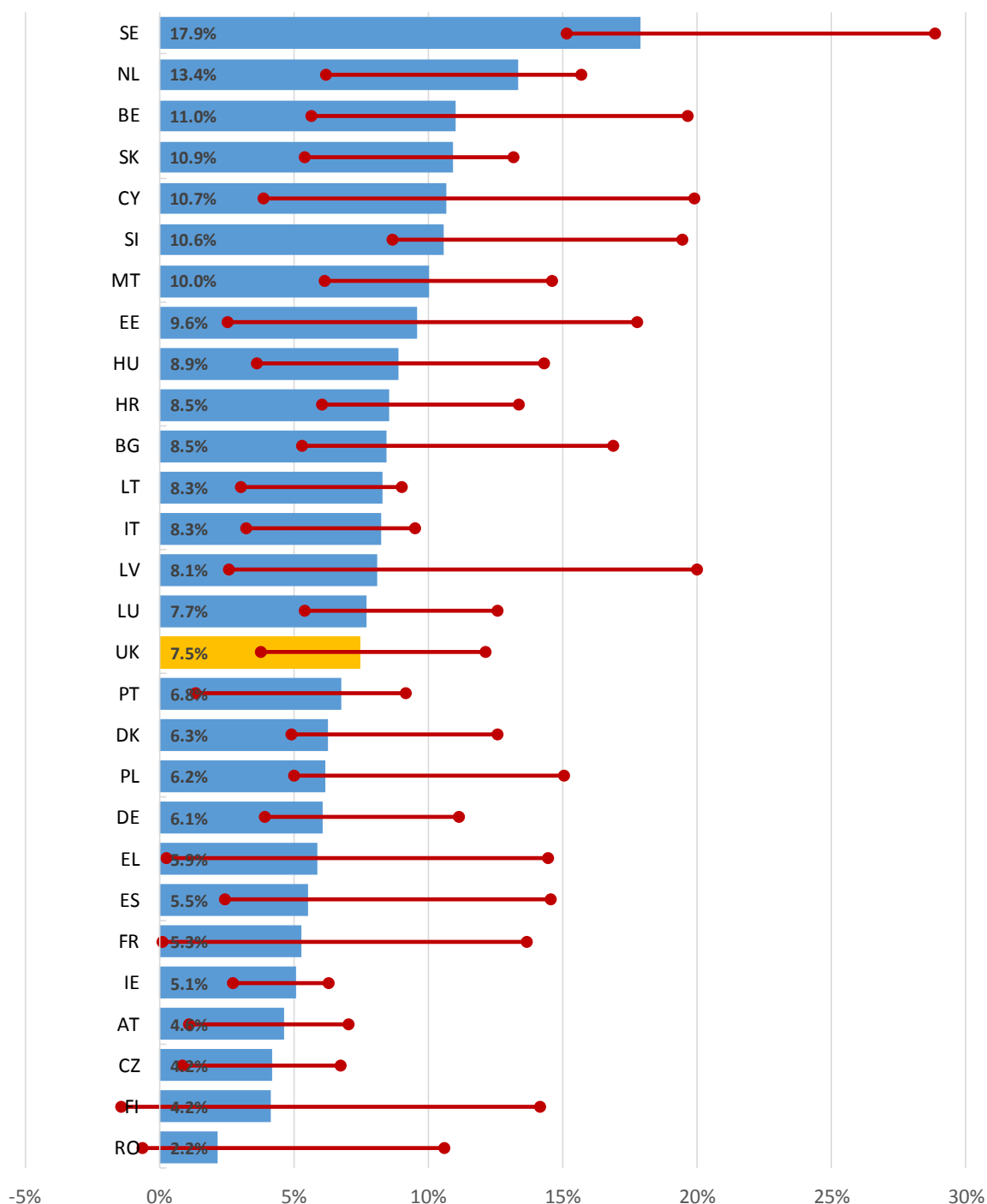
Source: European Data Market Monitoring Tool, IDC October 2016

According to the Baseline scenario, highest growth continues to be expected in Member States characterized by substantial ICT spending growth, such as Sweden, the

120

SMART 2013/0063 European Data Market

Netherlands and Belgium. As in our previous estimates, smaller economies like Slovakia, Cyprus, Slovenia, Malta and the Baltic countries, will rapidly catch up with respect to their more advanced neighbours and reach over-the-average growth rates by 2020 in all three scenarios. Some of these Member States, however, will continue to pay the price of their weak structural economic situation if the Challenge scenario holds true. Cyprus, Hungary, and the three Baltics, for example, would all exhibit growth rates well below the EU28 average by 2020 in this case. Greece will continue to struggle with a data market value under the EU average in all three scenarios at 2020 and will not be alone – Portugal, the Czech Republic and Romania will follow the same path.

7.3 Key Findings

The European data market continues to be an emerging phenomenon but its consolidation at both country and industry level remains undisputed. In fact, its overall value has shown a very healthy growth in 2016 and only minor fluctuations are to be expected in the 2020 forecasts according to the three different scenarios. The recent changes in the political landscape of the EU have certainly triggered some immediate reaction that has been accounted for in our new estimates. However, these reactions will gradually smooth out over the next couple of years.

- The European data market in the EU28 is estimated at EUR 54,351 million in 2015 and at EUR 59,539 million in 2016, thus exhibiting a solid year-on-year growth of 9.5%. This positive trend will continue throughout the next four years, although at different paces according to the selected scenarios, registering a 7.5% growth rate under the Baseline scenario, a 15.7% under the High Growth scenario and a 4.3% under the Challenge scenario.
- If we exclude the UK from our projected estimates, the value of the data market will grow at the same pace in the EU27, although slightly slower in both the Challenge scenario (with a CAGR of 4.1% vs. a CAGR of 4.3% in the EU28) and in the High-Growth scenario (with a CAGR of 14.9% vs. 15.7% in the EU28). As one of the most vibrant ICT-driven economies, the UK confirms its importance in an economy which is more and more driven by data and data-related products and services. As a result, its exclusion will have a negative (although minor) impact on the overall dynamics of the European data market.
- As a percentage of the total ICT spending in the EU28, the data market represents a share of 9.5% of the overall ICT market in the EU - against a share of 8.8% in 2015 – and is expected to improve under all of the three scenarios with respect to our previous estimates (14.8% under the Baseline scenario; 11.1% under the Challenge scenario and 14.5% in the High Growth scenario). The emerging nature of the data market makes sure that its growth will remain relatively unaffected by the variations in demand that are recently hitting the European ICT market as a whole. The overall ICT spending is therefore projected to increase at a more moderate pace in the years to come, while the data-related components of its spending (i.e. the data-driven technologies such as Big Data and analytics) will increase at a faster pace.
- At member state level, most of the leading EU economies will continue to grow steadily in 2016 with respect to 2015 – this is notably the case of Germany (growing 10% year-on-year), France (7.5% year-on-year) and Spain (4.1% year-on-year). Interestingly, also smaller and dynamic ICT-oriented economies such as Ireland (growing 18.8% year-on-year) and Denmark (14.2% year-on-year). Interestingly, less mature economies such as Croatia, the Czech Republic, Hungary, Romania, Slovakia and the Baltic countries will all witness a healthy growth in their data market

121

dynamics with growth rates between 11% and 17% year-on-year – a clear sign that the data market is consolidating and acquiring a self-governing dynamic.

- By industry, Europe's data market continues to be dominated by sectors that make a significant usage of data-related technologies, i.e.: manufacturing, finance, the area of professional services and the ICT sector. As in our previous analysis, the most dynamic growth by industry in 2016 is performed by vertical markets that still hold a relatively small size of the overall data market spending by industry. This proves that the data-related technologies are rapidly finding new ground in previously unchartered areas and are growing fast in sectors like home, healthcare activities, education and the public sector as whole.

- As in the first and second rounds of measurements of the European Data Market Monitoring Tool, our forecast for the period 2015-2020 under all three scenarios indicates that the data market growth will continue to be closely interconnected with the Member States' relative economic strength and the size of their ICT market. As a result, the U.K., Germany, France and Italy will continue to hold the lion's share of the data market under the three scenarios. In particular, countries with a relatively high percentage of data market over their total ICT spending in the period 2015–2016 will keep performing very well during the years 2017 to 2020.

- In 2020, the highest growth under the Baseline scenario is expected in Member States characterized by substantial ICT spending growth, such as Sweden, the Netherlands and Belgium. As in our previous estimates, smaller economies like Slovakia, Cyprus, Slovenia, Malta and the Baltic countries, will rapidly catch up with respect to their more advanced neighbours and reach over-the-average growth rates by 2020 in all three scenarios. Some of these Member States, however, will continue to pay the price of their weak structural economic situation if the Challenge scenario holds true. Cyprus, Hungary, and the three Baltics, for example, would all exhibit growth rates well below the EU28 average by 2020 in this case. Greece will continue to struggle with a data market value under the EU average in all three scenarios at 2020 and will not be alone – Portugal, the Czech Republic and Romania will follow the same path.

8. MEASURING THE DATA ECONOMY

8.1 Definition

*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.*

Indicator 4.2 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 will then be 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.
2. **The indirect impacts:** indirect impacts are all the impacts which take place in other industries related to the considered industry, in our case the data industry. 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 depending on the adoption of data by the downstream industries, i.e. the data users as classified in a selected number of industries as explained in the Methodology Report (Annex 3). 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 created by additional payment of wages to staff in the data industry and its direct supply chain. A proportion of this will be spent on consumer goods and services. This leads to further business growth throughout the EU economy. The IDC model quantifies the economic impacts on other industries as data workers spend their earnings. The additional consumption of data workers and of data companies' suppliers will in fact

123

SMART 2013/0063 European Data Market

support economic activity in various industries such as retail, consumer goods, banks, entertainment, etc.

As in the previous report, our 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.

8.1.1 Updating the Value of the Data Economy for the year 2016

The benefits related to the use of data products and services have been explored in our survey conducted between the end of the year 2014 and the 2015. In the survey, we asked about the benefits typology with reference to the year 2014 and about the benefits in quantitative terms for the year 2014 and 2015. We have assumed that there were substantially no changes in the kind of benefits the companies were having from the year 2014 to the year 2015. Besides, the survey asked about quantitative benefits for the year 2014, while for the year 2015 we asked what was expected in terms of quantitative benefits. For the estimates of the data products and services impacts we used both the data provided by the survey: the impacts registered by companies in 2014 and the benefits expected for the year 2015. We have assumed that the expectations of the companies were correct and matched the actual benefits. We believe by the way that this is a reasonable assumption since the market did not change substantially or in an unexpected way from 2014 to 2016.

To finalize our estimates of the impacts in 2016, the study team has also conducted additional desk research on a series of different internal and external sources, among which:

- 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: June 2016)
- Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: June 2016);
- IDC's detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 through 2016 (Second Quarter of the year);
- IDC Worldwide Black Book (Standard Edition), quarterly updates form the years 2014 through 2016. 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
- IDC European Vertical Markets Survey, 2015: More Western European SMBs Will Invest in Software Solutions Beyond Maintenance, July 2015
- IMF World Economic Outlook (WEO) Database, October 2016
- Consensus Forecasts, Consensus Economics, monthly updates, July – October 2016
- OECD Digital Economy Papers, among which: OECD (2014), Measuring the Digital Economy: A New Perspective; OECD Publishing.
- ILOSTAT (International Labour Organization) Statistics and Databases (2016)
- EUROSTAT Tertiary Education Statistics (Last update: December 2015 – Upcoming November 2016).
- European Data Science Academy (EDSA) project deliverables and publications (July 2015).

- IDC updated the indicators on data market, data companies, data companies' revenues, and the data economy by leveraging a variety of inputs, including but not limited to:
- 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: June 2016)
- Eurostat annual structural business statistics with a breakdown by size-class are the main source of data for an analysis of SMEs (latest update: June 2016);
- IDC's detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 through 2016 (Second Quarter of the year);
- IDC Worldwide Black Book (Standard Edition), quarterly updates form the years 2014 through 2016. 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
- IDC European Vertical Markets Survey, 2015: More Western European SMBs Will Invest in Software Solutions Beyond Maintenance, July 2015
- IMF World Economic Outlook (WEO) Database, October 2016
- Consensus Forecasts, Consensus Economics, monthly updates, July – October 2016
- The same sources were used to estimate the indicators on Data Workers and Data Workers' Skill Gap. For these two specific indicators, however, the study team also leveraged the following sources:
- OECD Digital Economy Papers, among which: OECD (2014), Measuring the Digital Economy: A New Perspective; OECD Publishing.
- ILOSTAT (International Labour Organization) Statistics and Databases (2016)
- EUROSTAT Tertiary Education Statistics (Last update: December 2015 – Upcoming November 2016).
- European Data Science Academy (EDSA) project deliverables and publications (July 2015).
- IDC estimated the indicator on the Citizens' Reliance on the Data Market by using the above-mentioned sources plus 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, June 2016.
- IDC European Quarterly Wearables Tracker Results: Western Europe 2Q16 Analysis, September 2016
- IDC FutureScape: Worldwide Wearables 2016 Predictions, November 2016.
- The Brexit Impact on IT Spend in the U.K. and Western Europe: A Scenario Analysis: An IDC Insights, July 2016.

8.2 Indicator 4.2 and 4.3: Value of the Data Economy and Incidence of the Data Economy on GDP

N.	Name	Description	2015	2016	Growth rate 2016/2015	Impact on GDP 2016
4.2	Value of the Data Economy EU28	Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy	285,633	299,989	5.03%	1.99%
4.2	Value of the Data Economy EU27	Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy	228,985	238,699	4.24%	1.92%

Indicator 4.2 — Value and Impact on GDP of the Data Economy in 2020, Three Scenarios (€ Million; %)								
N.	Name	Description	2020 Challenge (000)	2020 Baseline (000)	2020 High Growth (000)	Impact on GDP 2020 Challenge	Impact on GDP 2020 Baseline	Impact on GDP 2020 High-Growth
4.1	Value of the Data Economy EU28	Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy	361,325	430,259	739,383	2.30%	2.48%	4.00%
4.1	Value of the Data Economy EU27	Value of the Data Market plus estimate of direct, indirect and induced impacts on the economy	289,851	349,228	572,588	2.22%	2.44%	3.76%

Source: European Data Market Monitoring Tool, IDC October 2016

The overall value of the data economy grew from the EUR 247 Billion in 2013 to the EUR 285 Billion in 2015, almost reaching EUR 300 Billion in 2016, marking a growth rate year-on-year of 5.03% in the period 2015-2016. According to our estimate the value of the data economy in 2016 was worth nearly 2% of the European GDP – a narrow increase with respect to the previous year where the overall data economy represented a share of 1.94% of the EU28's GDP.

There are 15 countries with an impact of the data economy over the EU average. Among these countries we find, not surprisingly, Germany, the UK, Finland as well as countries like Estonia and Finland.

Major impacts can be found in Member States where:

- Data products and services have a high penetration in the domestic economy so that the benefits from using data and introducing innovation are widespread.
- Companies adopting data products and services have greater benefits than companies in other countries.

What is more, the introduction and diffusion of new technologies - like data technologies – is likely to provide immediate and sizeable impacts to those economies that are relatively new adopters of innovation: this is notably the case of Malta, and Cyprus.

As a whole, the total impacts of the data products and services are relevant in terms of GDP and correspond to the benefits deployed by the most innovative products.

Total indirect impacts are the most important component of the data economy as they involve backward and forward industries. The forward indirect impacts are vital as they involve all data user industries, which, as explained in the Methodology Report, are a significant part of the overall industry. The indirect impacts represent in fact 56% of the total impacts, and in some countries 65% - a portion that slightly increased in the last year compared to previous periods. The impacts of the diffusion are not producing major changes in terms of impacts. However, we should not conclude that the diffusion of data products and services is not having a significant impact in the short term, as it represents a limited period to measure the full potential and impacts for an emergent technology.

Figure 40 Value of the data economy and impact of the data economy by MS as a % of GDP, 2016
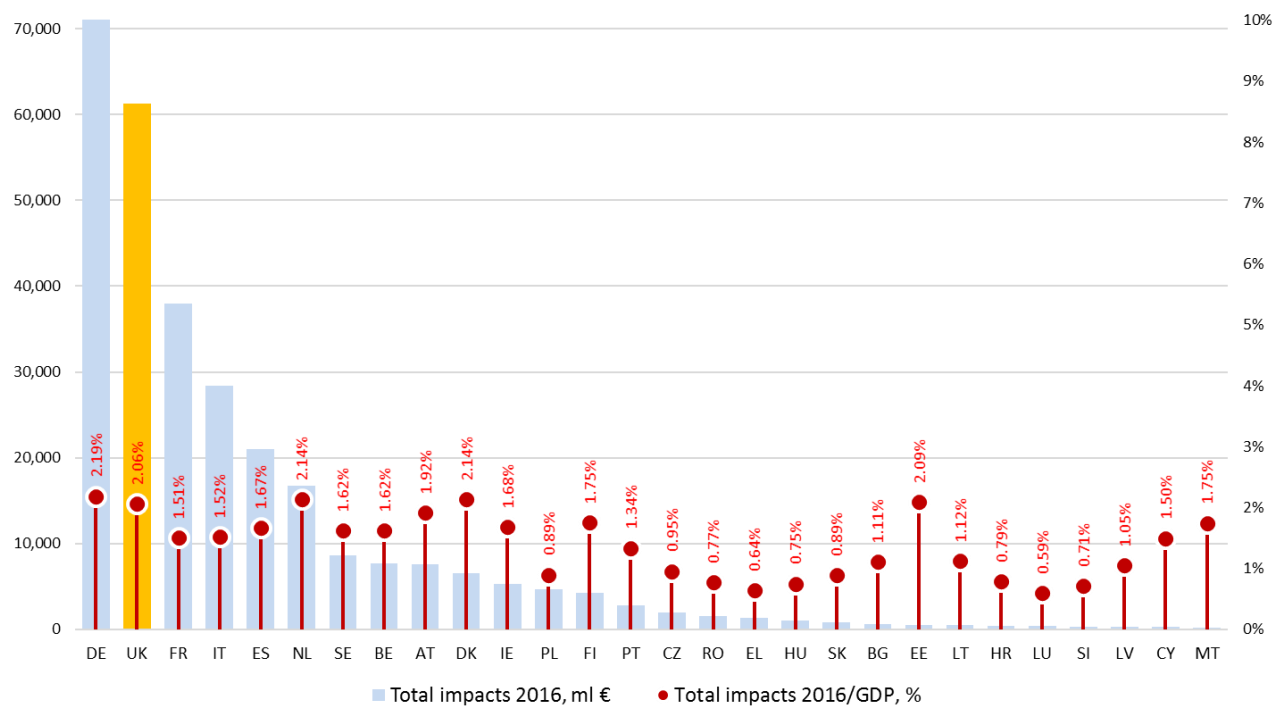
Source: European Data Market Monitoring Tool, IDC October 2016

The indirect impacts are strongly correlated with the data market: where the demand is more developed the impacts produced through users are more relevant in terms of absolute value. For this reason, that the total impacts on GDP by Member State present an extremely varied picture as they fluctuate from 0.6% in countries like Romania to about 2.1% in more advanced economies such as Germany, Denmark. Furthermore, some smaller Member States such as Malta, Cyprus, Estonia, Latvia, and also above-average total impacts of the data economy on GDP, thus mirroring their performance in

SMART 2013/0063 European Data Market

other areas measured in by other indicators, primarily by the value of the data market. We should be aware that the growth rate for the accumulated impacts may not be as fast as the growth rate for the data market because of the following reasons:

- The diffusion of innovation evens the competitive advantages of the adopters, and prices of the new technologies and products tend to decrease. This has to be considered an advantage for consumers, also in terms of quality of processes and products, but it does not necessarily translate into an increase in revenues and GDP.
- Data should be considered to be a multipurpose innovation, which in part may be a substitute for other products and technologies so that the net effects may decrease over time.

Also, the overall impacts of the data economy very much depend on penetration rates of data products and services. Since these rates are nearly stable over the observed years, so are the impacts relatively stable. What really changes the incidence of impacts is a widespread diffusion of data-driven innovation throughout the whole the economic system. As explained, we are currently in the first stages of the technology diffusion curve with penetration rates logically following a rather moderate trend. The overall induced impacts, which were worth over EUR 63 Billion in 2013, EUR 66 Billion in 2014 and almost EUR 73 Billion in 2015, made a 9.4% growth in the period 2015-2016 reaching nearly EUR 80 Billion in 2016. Taking into account that the estimate responds to a conservative approach, this represents a bit more than 0.5% of European GDP in 2016. This is based on two main reasons:

- Data workers do not necessarily correspond to new jobs: therefore, all the data workers are not necessarily going to earn additional wages;
- During an economic crisis, the propensity to save increases, which may extend the typical time-lag of such impacts (they tend to occur later than the market launch of new products and services).

Figure 41 Impacts of the Data Economy by MS as a % of total impacts, 2016
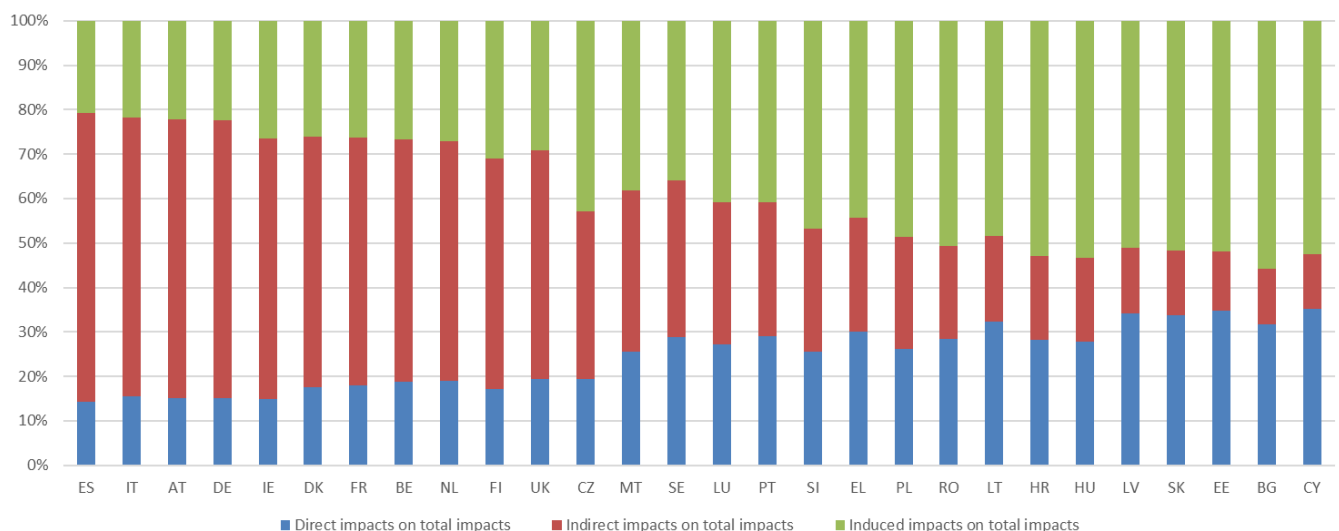
Source: European Data Market Monitoring Tool, IDC October 2016

Overall, the indirect impacts, including forwards and backwards, are the most relevant impacts, although this is not true in all countries. The countries where the indirect

SMART 2013/0063 European Data Market

impacts are more relevant are the one leading the data market. The estimates show the larger shares of indirect impacts in countries like Spain, Italy, Austria and Germany. On the opposite, in Estonia, Bulgaria and Cyprus the larger amounts are represented by the induced impacts on the overall economy.

8.3 Analysis and Interpretation

The amount of data generated, collected, managed and elaborated through data analytics tools is exponentially increasing. Such an amount of data and its management are dramatically changing the knowledge process creation of social and economic systems and the decision making processes as well. Analysts underlined that the new decision-making processes act as a rationalization and optimization 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 view 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 specialized data analytic 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. This is by the way out of our field of analysis, but since we are dealing with data impacts we are going to present the most relevant ways in which data provide benefits to organisations (McKinsey, 2012, OECD, xxx). 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 allow 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. We should stress here that these are also the more significant impacts in terms of GDP.

Figure 42 Data economy: Distribution of impacts as a % of total impacts, 2016

■ Direct impacts   ■ Backward indirect impacts   ■ Forward indirect impacts   ■ Induced impacts
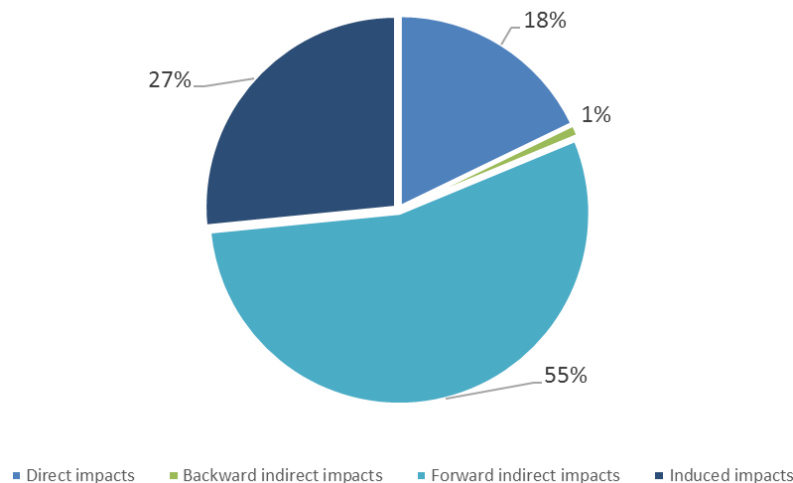
Source: European Data Market Monitoring Tool, IDC October 2016

8.4 Indicator 4.2: Data Economy Forecasts

We estimate again the impacts of the data economy for all the 28 EU Member States again in 2016. As stated in the assumptions, we assumed that a number of countries are and will have similar impacts as, in average, the main countries. For the forecasts, we also assumed that:

- the impacts are positively correlated with the increase of the market in general
- the impacts are going to be strongly related with the penetration rate of the data products and services within the economic system.

The total indirect impacts (including backward and forward industries) of the data industry itself, are the most relevant effects, followed by the induced impacts.

To analyse the impacts at 2020, we estimated the impacts within three different scenarios: Challenge, Baseline and High Growth. The scenarios are fully explained in Chapter 3 of the current report. The scenarios are based on different macro-economic trends. In the Challenge scenario, Europe is going to recover slowly, and the GDP growth and the ICT spending are going to increase gradually: this will limit the innovation propensity of companies and, therefore, the adoption and diffusion of new products and services.

Figure 43 Impacts of the Data Economy by MS as a % of total impacts, 2020 Challenge Scenario

SMART 2013/0063 European Data Market
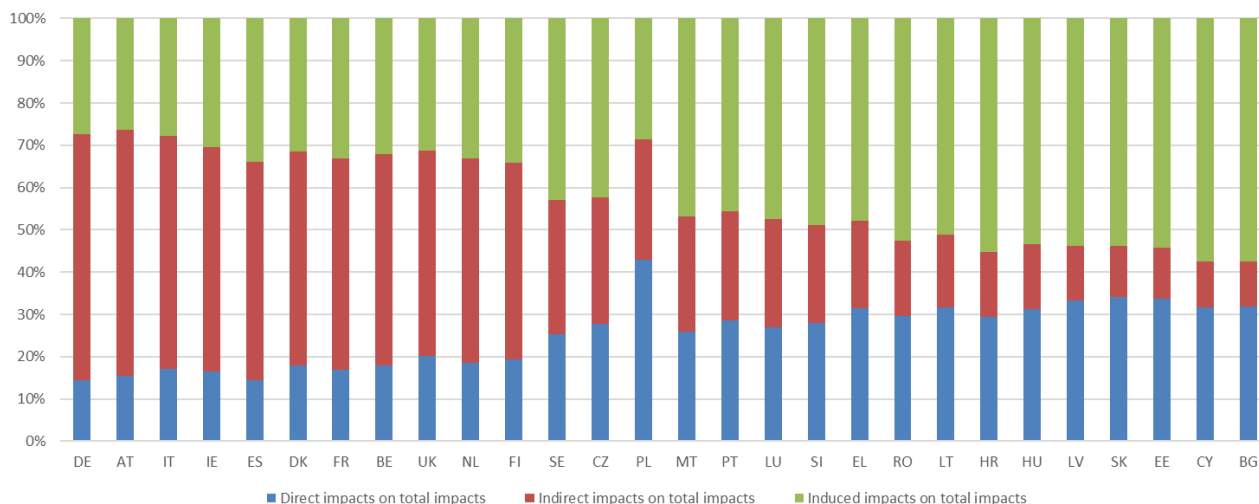
Source: European Data Market Monitoring Tool, IDC October 2016

In the Baseline scenario, on the opposite, Europe will recover from the crisis showing a GDP trend similar to the pre-crisis years and an increasing ICT spending. Such positive macro-economic trends will support investments for innovation and, therefore, the adoption of data products and services. The Baseline scenario highlights the impacts of data products and services with a positive economic trend.

Figure 44 Impacts of the Data Economy by MS as a % of total impacts, 2020 Baseline Scenario
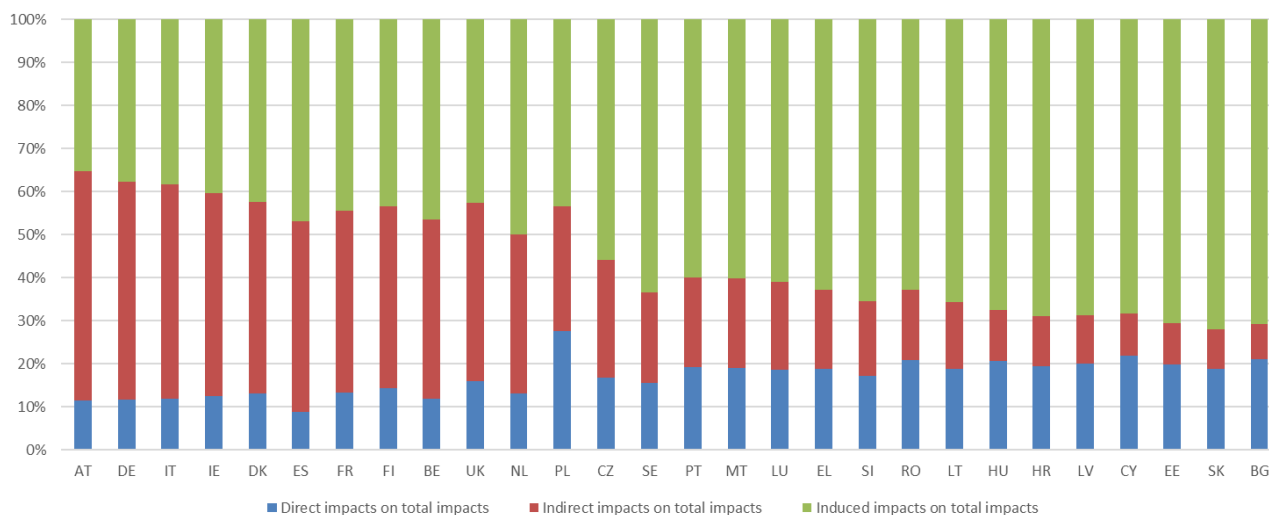
Source: European Data Market Monitoring Tool, IDC October 2016

The High Growth scenario foresees macro-economic trends similar to the Baseline scenario. Nevertheless, the ICT technology push will support companies in daring investments helping cost savings and new benefits. Users will demand data products and services more than they did into the Baseline scenario thanks to an awareness effect about the achievable benefits of the data innovation. As a consequence, in the High Growth scenario, the demand of data products and services from companies will grow,

132

and the penetration rate in the user industry will be twice the penetration rate of the Baseline scenario.

The High Growth scenario, therefore, explores the impacts of data products and services with a GDP equal to the one in the Baseline scenario, but with an increased penetration of data services and products.

Figure 45 Impacts of the Data Economy by MS as a % of total impacts, 2020 High Growth Scenario
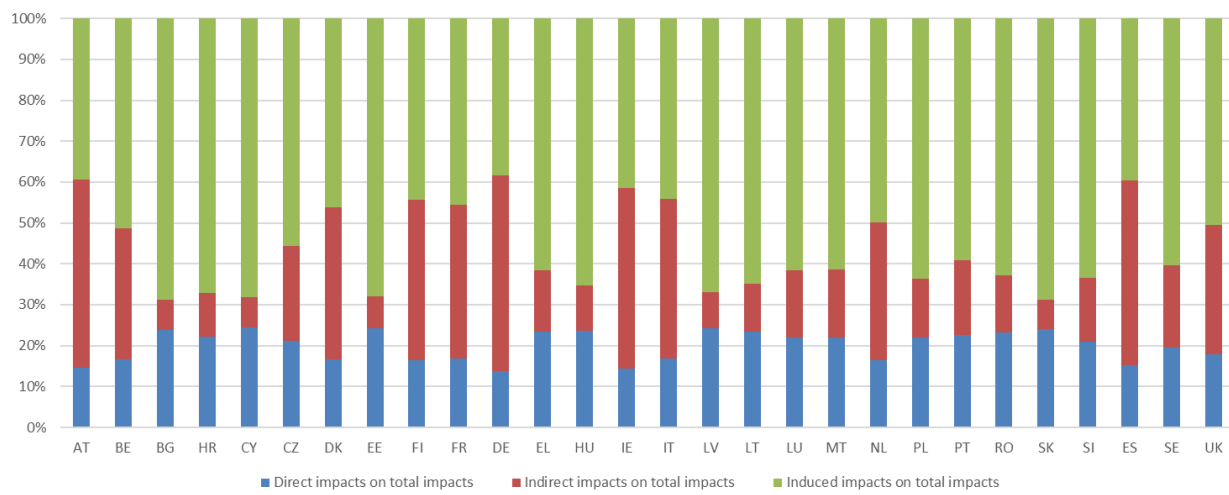
Source: European Data Market Monitoring Tool, IDC October 2016

The total impacts of data products and services at 2020 across the countries may vary, in the three scenarios, from 2.3% of the GDP in the Challenge scenario to nearly 2.5% in the Baseline scenario and to 4% in the High Growth scenario.

The forward indirect impacts are confirmed to be the most important. These impacts are all the impacts deployed through the user industries. In 2020, and within the Baseline scenario, the forward indirect impacts will be supported by an increasing penetration rate of data as well as by increasing benefits due to gains in efficiency and competitiveness in the user companies.

We have confirmed a conservative estimate of the induced impact because, as explained in the data worker chapter, data workers do not correspond to net job creation and a share of the data workers are people already working in the companies, dealing with management or ICT.

In the Challenge scenario, the forecast value of the data economy will represent about 2.3% of GDP of the EU28 GDP. This narrow difference between the Baseline scenario (2.48% of GDP) and the Challenge scenario depends on the fact that the penetration rate between the two scenarios is narrow. This means that the data diffusion is at a stage where the diffusion of such new products and services will occur in both scenarios. We need to note here that our Challenge scenario is a moderately pessimistic one which is not going to stop the innovation diffusion especially where innovation is able to provide benefits in terms of more efficiency and competitiveness.

Besides, the High Growth scenario is characterised by a much higher penetration of data into the user industry. The High Growth scenario shows an impact on GDP of 4%. This

133

relevant impact is mainly driven by the forward indirect impact as it was in the other two scenarios. Nevertheless, in this case, the forward indirect impacts represent 67% of the total impact, more than it was into the other scenarios. This effect can easily be explained: in this scenario, the penetration rate of data services and products into the user industry is much faster than it was into the other scenarios; this is the reason why in the High Growth scenario the impacts achieved through the users are more relevant than they were in the other scenarios. The multiplier effect of data product and services is higher when, with equal macroeconomic conditions, the penetration in the user industry is higher.

On the overall, we confirm the forecasts already presented last year, although the current ones are a bit more positive, due to the faster increase forecasted for the data market.

When looking at the impacts at country level, we observe that in the baseline scenario, there are 11 countries with a forecasted impact on GDP higher than the average forecasted for the EU, while the other 17 countries register a lower incidence of the impacts on GDP. The countries with the higher share of impacts on GDP are the countries where the demand for data products and services will grow with a fast trend. Where the penetration rate is faster, the impacts on GDP will be even more relevant.

We should here remind that the market can increase because of a higher intensity of demand (the same companies demand more products) or because of a faster penetration (more companies demand more data products). Based on our analysis, it seems that, at the same level of the total demand, a higher penetration rate will produce higher impacts, measured as a % of the GDP.

8.5 Key Findings

- The economic impact analysis is an effective tool for the scrutiny of the impacts of a multipurpose and widespread innovation such as the data products and services. This kind of approach helps subdivide the impacts in order to better understand the source of such impacts, and whether they can be considered new additional impacts on the economic system. An in-depth and reliable analysis focusing on impacts should be based on specific field research, but the survey provided interesting insights about the impacts gathered by the companies adopting data products and services. A significant share of the companies is starting to see benefits in terms of additional revenues.

- The latest estimate of the total impacts for 2016 accounts for 1.99% of the EU GDP, corresponding to almost 300,000 Million. In 2013 the total impacts estimated for data products and services represented EUR 246,840 Million, which is equivalent to 1.83% of the EU GDP; in 2014 EUR 257,589 Million, which is equivalent to 1.85% of EU GDP. In 2015, we estimated the same indicator to be at EUR 285,633 Million, representing 1.94%

The qualitative analysis of the impacts, based on the survey results (specifically for the years 2013 and 2014), as well as the IDC cases-studies and the stories collected for this study, shows that the main impacts in terms of GDP are the forward indirect impacts, which are the impacts deployed on the user industries. The type of innovation we are dealing with and its adoption process also suggest that this impact is the "actual" new impact emerging in the economic system.

- The impacts gathered by the user industries (forward indirect impacts) represented almost 56% of the total impacts in 2013, approximately 54% in 2014 and remain substantial through 2016, corresponding to more than 1% of total EU GDP in both 2015 and 2016 (1.08% and 1.09% of total GDP respectively).

When compared with the total impacts in 2013, we can see that the total impact is not increasing very fast and significantly year on year. This is quite normal and in line with the overall impacts of innovations. First of all, because the penetration of the data products and services is not so fast and also because the growth rate for the accumulated impacts may not be as fast as the growth rate for the data market (see 8.1).

The user industries are starting to see the quantitative benefits from the use of data. These benefits on the user industry translate into revenue increases. In turn these lead to GDP increases in the region of 1% for the EU.

- The scenarios at 2020 show that a high penetration in the user industry produces relevant and fast impacts in terms of GDP.

The High Growth scenario, under similar macroeconomic conditions, produces relevant impacts on the user industries. A fast penetration of data products and services produces relevant effects in terms of GDP though the benefits achieved by the user industry. This means that policies that leverage on increasing demand for products and services may provide relevant impacts.

9. THE DATA MARKET AND STORIES: A COMPLEMENTARY APPROACH

9.1 Background and Objectives

While the European Data Market Monitoring Tool remains the pivotal instrument around which quantitative evidence is collected and analysed, descriptive evidence is also required to produce an all-encompassing assessment of the data market and the data economy. In fact, figures alone cannot entirely catch the complexity of such an emerging phenomenon. Furthermore, the data market rests upon the concept of process innovation, whose impacts on the economy and markets tend to materialize after quite some time from its appearance and is not immediately translated into data and figures. Policy makers are therefore in need of additional sources of qualitative evidence to timely apprehend the effects of innovation on the data market, as well as the impacts of the latter on the overall economy.

In this study we presented the results of a series of descriptive stories providing a complementary view to the one offered by the Monitoring Tool and adding fresh, real-life information around the quantitative indicators. The aim of these stories was to investigate the aspects that characterise the creation of the data economy in general, and the data market in particular, and that are not captured through the measurement of the indicators under the European Data Market Monitoring Tool. By focusing on specific issues and aspects of the data market (such as the effective implementation of new data-driven products and/or services; the presence of efficiency gains obtained through the use of data-related product and services; the ability to improve decision-making processes at both organisational and societal levels, and more), the stories offered an initial, indicative "catalogue" of good practices of what is happening in the data market today in Europe and what is likely to affect the development of the data economy over the next few years.

9.2 The Stories So Far: An Overview

At its completion, this study will produce a total of twelve different stories investigating the use and impacts of data-related technologies and data-driven innovation across the European data economy. The previous reports have shown the results from the first eight stories covering the public sector (with specific reference to the central and local government spaces) as well as a wide portion of the private sector: from the retail to the utilities industries, from the manufacturing to the banking industries. The present report integrates the results of three additional stories focusing on the role of data with specific reference to the necessary platforms and architectures that are currently being developed to ensure a smooth data sharing and data exchange, as well as to the skills that more and more European companies in varied industry sectors are trying to promote internally or acquire externally in order to sustain the digital transformation of their organizations.

9.2.1   The Starting Point: Innovation and Competitive Challenges

Two phenomena appear to be common to all the stories developed so far:

- Data-related technologies are emerging but are not completely new in the industries and in the sectors under consideration. In other words: technologies such as Big Data and analytics software are certainly innovative but can rely on long-established practices of heavy data usage in most of the industries that we have investigated so far.
- Data-related technologies are proving to be effective against the damages that the economic crisis has caused in the industries and sectors under consideration. The worsening economic environment has accelerated the quest for greater competitiveness and efficiency, and data-related technologies are central to this.

A few examples from the stories that we have developed can be used to further clarify these two fundamental elements.

The Retail Industry

In the retail industry, a new generation of technology-savvy consumers is rapidly emerging, forcing retailers to rethink their marketing and communication strategies. For example, consumers can look for the desired article on the Internet, try it in a store, and buy it from a mobile device at the best price that they can find. To counter this, retailers can now count on a wide range of data-driven demand forecasting and advanced pricing analytics tools that can process a very large amount of data in real time and produce very accurate forecast models. These models, in turn, can provide retailers with significant support to improve brand performance, drive customer loyalty, adjust pricing, and improve customer satisfaction.

The Government Sector

The government sector is one of the biggest "data producers" in the economy: this is valid in the U.S. as well as for most of the European Member States. On both sides of the Atlantic, austerity measures, and hence cuts in government spending, are pushing government agencies to identify opportunities for more efficient operations and interventions. Big Data and analytics technologies can have a positive impact at different levels in the public sector: at the level of service provision, at the level of public policy making, and at the level of policy strategy. Our research shows that data-driven technologies can positively affect public service provision in Europe — successful applications are particularly visible in transport, crime prevention, tax collection, garbage collection, and disaster management, to name a few. At policy and strategy level, EU governments are not as advanced as others but there are some exceptions, such as the U.K., which is very active in open government data initiatives, and the Netherlands, which is one of the main players in the use of data for anti-money-laundering purposes.

The Manufacturing Industry

Today's data availability (and associated data technologies) also represent a great opportunity for the European manufacturing industry. Faced with one of the worst economic crises since the Second World War, European manufacturers are increasingly turning more and more to data-driven solutions and applications to optimise production processes and reduce operating costs. As in the retail industry, however, the extensive availability of data poses a growing threat as buyers become increasingly knowledgeable about products, prices, and other key features. As a result, Europe's manufacturers are putting in place an ever closer collaboration with technology companies to implement

137

new IT architectures enabled by Big Data, business analytics, and other fundamental data-driven technologies such as the Internet of Things and cloud computing. This provides them with new ways of analysing and interpreting the information that, most of the time, they possess inside their organisation in the form of structured or unstructured data.

The Utilities Sector

European utilities are in the midst of an evolutionary transformation that started in the early 2000s and will peak in the course of the next decade with the mass rollout of gas and electricity meters in most of the EU Member States. This transformation entails a lot of risk for Europe's utilities players, such as the need to optimise existing network capacity, integrate a growing amount of distributed renewables energy resources, increased demand for electric mobility, and, more generally, an overall power shift to the downstream segments of the utilities' value chain which could cause an exponential generation of data at a more decentralised level. Our research in the utilities sector in Europe shows that data-driven applications and solutions are already having a beneficial effect on a wide range of processes and services: they help maximise process efficiency and productivity of service teams in the field (field service optimization), improve asset demand forecasting, production planning, energy and water usage and supply, or better analyse and predict outages at central generation/abstraction plants (asset operations). All in all, our stories confirm that European utilities are fully exploiting the advantages offered by data-driven technologies and are becoming *the* data-driven industry in Europe.

The Banking Industry

The European banking industry has been hit hard over the past few years. However, after a long period of restructuring, banks in the EU have restored their financial position and resumed investments in IT. The bulk of these investments are devoted to Big Data and Analytics technologies (BDA), which are helping the European banking industry stand firm in the face of pressures that range from the need to restore their brand reputation, renew their customer centricity, ensure full compliance to a rapidly changing regulatory framework, and refocus on core banking activities. Our research shows that BDA implementations are becoming more and more popular among a growing number of banking institutions and are bringing tangible benefits in several areas of the industry value chain: operations can be successfully improved through the deployment of a new generation of analytics providing enhanced visibility into business processes and events, for example; customer engagement can also be positively affected by Big Data and analytics technologies applied to marketing and customer relationship management solutions allowing banks to target the right customers with the right products and at the right time; risk management processes can also be greatly enhanced through more effective use of data which can help to detect suspicious activities and reduce banks' exposure to risk.

Precision Agriculture

Agriculture is undergoing an in-depth transformation, and is moving fast from a traditional sector to a strongly IT- and business-oriented market. The trigger of this revolution is represented by a broad set of new technologies, grouped under the term "precision agriculture": smart sensors, drones, high precision positioning systems, smart applications and devices combined with high-tech engineering. In addition to significant

gains in efficiency, productivity, cost management and cost reduction, our research shows that data-driven technologies are helping the sector face some of its newer challenges: smart farming technology today is addressing pressing concerns, such as labor shortages and environmental factors. Precision agriculture is also about developing new skills and building partnerships. As the worsening economic environment has accelerated the quest for greater competitiveness, precision agriculture represents yet another example of how data-related technologies can help businesses in devising not only new products and services but also new forms of partnership and business collaboration.

9.2.2   A lateral view: The Issue of Data Ownership

In parallel with the role of data-related technologies in supporting the data market and exerting positive impacts to the European data economy in general, the study team also addressed a key element directing affecting the very availability and usage of data – the concept of data ownership and its related issues. In fact, the way data are made available and the extent to which data are flowing across sectors and organizations, play a fundamental role in sustaining and developing the emergence of a European data-driven economy. In defining and specifying the rights to create, edit, modify, share and restrict access to data, data ownership becomes a pivotal factor affecting a growing number of potential data users and an increasing range of data-related activities.

Our analysis shows that European businesses find it difficult to adopt a viable, shared definition of data ownership and resort to existing intellectual property rights' (IPR) regimes, or current database rights' systems, to safely exchange data. As a consequence, most of the stakeholders that have participated to this study do not exercise any pressing request for new, data-related contractual arrangements or alternative types of regulatory regimes specifically oriented to manage data ownership and are content with the existing contractual forms in use. At the same time, our investigation reveals that Europe's rapidly evolving data market is constantly putting forward new business models and that data stakeholders in Europe may well benefit from ad-hoc guidelines to adjust the existing contracts to these emerging data-based business models. If more and more European companies do embrace innovative data-based business models, no evidence emerges yet as to the existence of a well-functioning, shared and recognized data-pricing mechanism. Indeed, data are often exchanged "in bundles" with other services so that their value is inevitably included in the overall price that businesses apply to the bundled service.

The way data ownership and data access are managed and regulated can directly affect the functioning of the data market. Companies having a high concentration, or accessing huge amounts of data, could easily incur in situations of market asymmetry, which – in turn – may result in different forms of market distortion. In our case studies we found no significant evidence of severe market abuse as the current level of data exchange and data re-use does not seem to cause stark hindrances to the overall market efficiency, at least at this stage of the process. As a result, data ownership could be considered within the broader framework of growth, innovation and competition policies and not seen simply as a contractual issue or a legal matter. This is not to say that a certain number of guidelines, as well as new types of model contracts, could be fruitfully developed by the industry to help data-stakeholders come to terms with emerging business models and new business cases.

9.2.3 Enhancing Data Sharing and Data Exchange: The role of Data Marketplaces and Industry Data Platforms

The recent developments of the so called "3rd Platform of computing"[18] have significantly accelerated the exchange of electronic data giving birth to the phenomenon of (electronic) data marketplaces. In their simplest form, data marketplaces are online stores where people can buy and sell data, but they are evolving towards a more sophisticated intermediary role, potentially central to the emerging data ecosystem. They can maximize the value of data by facilitating the exploitation and re-use of proprietary data as well as its integration with open data, by providing a platform for data holders, service providers and data users to work together, and by building trust in the data value chain, thanks to their role as independent third party. By doing this, data marketplaces play the role of multi-sided digital market platforms, where the value of the products exchanged is multiplied by the interaction of the platform users. Data marketplace platforms of this kind can respond to a variety of emerging data market needs, by improving the efficiency of the data value chain, reducing transaction costs, providing a platform for sharing and re-using data sets, and solving data interoperability, privacy and security problems on behalf of data holders who may not have the necessary skills.

However, this is only a nascent, although fast developing, phenomenon. In today's data market we have identified several categories of suppliers who are increasingly providing data marketplace services, but most of them are still closer to the simple model of the data store than to the sophisticated, multi-sided platform model. Furthermore, most of these companies are almost all owned by large technology providers, the majority of which are US-owned. Even companies of European origin, like Swedish Qlik, tend to move to the US to incorporate, find easier access to capital and to the richest IT customer base in the world. However, these suppliers do not completely satisfy the needs of European data holders as they do not provide yet the interoperability, security, privacy and legal compliance services required in the EU. Also, the offering in terms of datasets is still limited and has not yet been able to mobilize the majority of EU data holders. As a result, a considerable emerging market opportunity is starting to materialize in Europe and European entrepreneurs are starting to act consequently and founding several EU start-ups which are developing specific data marketplace offerings. Self-defined data marketplaces represent 3% to 3.5% of the more than 450 actors collected by the European data landscape community launched in the context of the European Data Market study; they are a very small group of players compared to analytics, data holders or ICT enablers companies - an empirical indication that data marketplaces represent an emerging and very early stage phenomenon in Europe. The potential demand related to data market places is high and could be satisfied by European actors, leveraging industry-specific databases, the trust and confidence of business partners, and the capability to satisfy EU standards of data quality, privacy protection and security. However, both the European suppliers and demand appear immature and still far from the ability to compete with leading US competitors.

Data marketplaces are not the only way organizations regularly exchange and share an increasing amount of data. Other forms of initiatives are emerging across the two sides of the Atlantic. Industrial data platforms, in particular, are rapidly becoming a trusted network for data access, transfer and usage, thus disrupting traditional value- and supply-chains and bringing together a wide array of disparate players from a multitude of

---

[18] http://www.idc.com/prodserv/3rd-platform/

sectors. However, at this initial stage, industry data platforms are not easy to define with their actual implementation displaying a diverse picture where two opposite models coexist. On the one hand, industry data platforms may take the form of open, multi-company-led environments that are conceived to meet the requirements of a wide community of industry users from different industry sectors; on the other hand, single-company initiatives are emerging where an individual organization (usually a prominent ICT player or a leading industry actor) establishes its own data platform and open it to other companies for commercial purposes and primarily within the boundaries of a specific industry sector.

In either case, industry data platforms are not only allowing data to be accessed, exchanged and used in a secure and trusted environment but they are also unleashing benefits that go beyond the mere technical and architectural aspects. Through the use of "connectors" and "application layers", they are also providing their participant companies with the possibility to develop customized applications and, as a result, an ever growing number of company-tailored digital services. This, in turn, generates economies of scale and networking effects that will be critical for the success of the overall industry digitization, whether at national, European or world level.

Companies are already starting to realize the advantages of this new model of exchanging data in terms of both productivity and efficiency improvements. IDC Manufacturing Insights, for example, foresees that by 2019, 75% of manufacturing value chains in EMEA will undergo an operating model transformation with digitally connected processes that will improve responsiveness and productivity by 15% on average. Again both the Boston Consulting Group and PwC believe that digitization and interconnection of products and services will also allow companies to again additional revenues of 2% to 3% per year on average. If confirmed, these gains could amount to approximately €30 billion per year for Germany's industry alone and to €130 billion per year for the European industry as a whole

9.2.4 Enhancing Data Sharing and Data Exchange: The Role of Technical Barriers

Data marketplaces and industry data platforms are only some of the most recent examples of technology developments enhancing data sharing within sectors and across sectors. However, several types of barriers may hinder the effectiveness of data transfer and reduce the value-add potential of data use, re-use and exchange. Business-driven or legal inhibitors (such as the deliberate choice to limit possible competition and the difficulty in valuing data) are not alone in potentially interfering with effective data sharing – technical barriers, and the costs associated with them, can also discourage companies and organizations to use and re-use each other data.

In theory, the lack of interoperability could represent one of the most important issues hampering data sharing between organizations and across sectors today. Indeed, our empirical research reveals that several solutions are already available to reduce the hindering power of insufficient interoperability. Narrow industry standards, for examples, or high level architecture standards can make data easily accessible and transferrable between operators in the same segment or at industry level. The increasing adoption of ad-hoc technologies such as APIs and SDK (Software Development Kit) may ensure third-party secure access to data and enable easier data sharing. What is more, machine learning technologies are expected to enhance data curation activities, thus facilitating the integration of data from different sources and organizations. Additional investments in human resources and data skills can also play a significant role: data curation usually

takes at least 50% of the data scientists' work in data marketplaces and data analytics companies but this and work is necessary to offer better data curation and integration services and, ultimately, to improve data sharing. In general, a growing towards greater openness and data sharing between organizations and across sectors is taking place nowadays in Europe. Market forces matter: data are becoming more valuable, the cost of data gathering decreases, and customers (especially business) expect greater accessibility of the data.

While the evolution towards greater data sharing is unanimously recognized, more needs to be done to increase the speed of this trend according to our interviewees. In fact, the market is already acting in response to users' demand for greater data accessibility, but the demand itself is not yet strong enough to accelerate the trend. Furthermore, universal interoperability is hardly achievable in full, nor is it fully desirable: different languages, protocols and models have the advantage of delivering context-specific data, and as such cannot be eliminated tout court.

Overall, interoperability issues appear to be more an additional cost (estimated at least at 50% of data scientists' work in data related projects) than a technical barrier in its own right. Similarly, the lack of industry standards may not necessarily represent a barrier but rather a business opportunity as an increasing number of data marketplaces, data companies and business analytics companies offer data curation as part of their services. Indeed, the idea of a neutral b2b data exchange could be beneficial but mainly as a way to raise the awareness of companies towards greater data sharing.

9.2.5   *Working with Data: Acquiring the necessary skills in the era of Digital Transformation*

According to IDC, 65% of large enterprises worldwide will have committed to becoming information-based companies in 2016, shifting the focus from resources, labor, and fixed capital to relationships, people, and intangible capital[19]. Understandably, new digital skills are required to reap the benefits of such shift but most of the EU companies and organizations do experience a skills mismatch issue: European employers are struggling recruiting people with the skills needed, unemployment remains at high levels and over 25% of young adult employees are overqualified for their jobs (Eurofound, 2014 and Cedefop 2015). This bleak picture is confirmed by the results of the European Data Market tool[20] : its Data Skills Gap indicator projects a potential supply-demand gap of approximately half a million data workers' jobs in the EU by 2020 under the baseline scenario. As a result, only 64% of European firms expect their digital transformation objectives for 2016 to be fully reached (European Commission, 2016; The Economist Intelligent Unit, 2016) and many companies are bound to consider new and more compelling initiatives to upskill, re-skill, or simply acquire newly hired employees with the necessary skills.

In terms of managing the needs for new digital skills, our analysis across European organizations highlights the importance of bringing about an all-encompassing process of

---

[19] IDC Digital Transformation FutureScape 2016: Worldwide Predictions, IDC 2016

[20] See: The European Data Market study (SMART 2013/0063): Second Interim Report, June 2016. Also in: http://new.datalandscape.eu/

cultural transformation in parallel with the digital transformation journey. In other words, the demand for highly sophisticated and specific data skills remains limited and not too difficult to meet through the acquisition of a small number of highly specialized data scientists. On the contrary, upskilling the digital skills of the overall existing workforce poses a much bigger challenge and requires a deeper change. When it comes to the training and upskilling of the workforce, our primary research shows a clear preference for internal training: it is less costly, it allows the process of digital innovation to be spread and steered by the company according to its chosen strategy and reduces the risks related to the exposure of business-sensitive information to external third-parties. As for the sourcing of data skills, European companies seem to favour the route of external acquisition through the hiring of young graduates with very specific data skills.

From the above, a few critical issues requiring policy-makers' attention emerge:

- The need to deal with employees with obsolete skills who are not willing or able to adapt to digital skills. Our research points to a widespread lack of awareness by enterprises and by employees about the depth of the cultural transformation currently underway.
- The need to improve the visibility of data skills recruitment challenges to help companies and organizations find the right people with the appropriate sophisticated data analytics skills;
- The need to accelerate the current industry-university collaboration to increase the number of data-related internships, their duration and their scope so to ensure that data competences are available and up-to-date.

9.2.6 Key Benefits and Impacts

The rapid emergence of a data economy, and the associated dissemination of a significant level of data-driven innovation across Europe, is accompanied (or enabled) by an ever advanced and widespread use of information technologies. Whether through the availability of enhanced connectivity infrastructures (personal area networks and wireless local area networks, in addition to wired connectivity), the growing role of cloud computing, or the skyrocketing number of smart devices, these technologies produce data in an amount, and at a speed, that were simply unimaginable a few years ago. At the same time, these technologies are creating cutting-edge solutions and applications that are able to analyse and make sense of an immense volume of data. Our empirical investigation on data-driven innovation through a number of European economic sectors endorses the hypothesis of a fast-moving adoption of sophisticated data-related technologies in the form of Big Data and advanced analytics solutions and applications (BDA). Our analysis also confirms that these solutions do bring about tangible benefits and visible impacts when properly applied throughout the different segments of the data value chain.

An all-inclusive summary of the whole benefits and the impacts produced by data-driven technologies is extremely challenging to draw. Nonetheless, the stories that we have developed offer sufficient evidence to attempt an initial classification. Our investigation into the nascent data market across Europe's public sector and manufacturing, retail, utilities, and banking industries has identified four main categories of impacts from which several benefits can be derived.

- **Increased revenues** — the use of BDA has a positive impact on customer and sales-related activities.

- **Reduced Costs** — BDA and data-driven technologies help reduce expenditures and allow for considerable cost saving measures.
- **Enhanced operational efficiency** — BDA solutions and applications generally contribute to reduce negative externalities affecting production and operations (i.e., downtime, rework, product returns, and time spent for maintenance and trouble-shooting).
- **Improved organisational and policy effectiveness** — enhanced, faster, and more accurate data usage and analysis improves an organisation's (whether a private company department or a public sector agency) ability to fulfil its institutional goals. Central and local government agencies, for example, can design and implement more effective policies, while private companies can reach their objectives in a more competitive and timely fashion.
- **Fostered entrepreneurship, new ventures and cross fertilization** – the ever increasing use of data and the widespread presence of new data-related technologies, produces a large and complex ecosystem involving growing numbers of various stakeholders. These may come from very different business segments and backgrounds, including sectors not traditionally familiar with advanced information technologies and with relatively low ICT penetration rates. In this context, BDA represent a new form of technology that finds easy adoption in both ICT-advanced and more traditional sector. What is more, data technologies can serve as an effective bridge between economic operators theoretically very different from one another, and allow new ways of collaboration and cross-fertilization of ideas that may turn into new form of businesses.

A summary of the main benefits and impacts, as well as a list of relevant quantified examples for these benefits, is outlined in Table 30.

Table 30 Impacts and Benefits from Data-Driven Technologies per Industry/Sector

Impacts	Benefits	Industry / Sector	Relevant Examples
Increased revenues	Increase in number of new or reactivated customers	Retail	Morrison's, a British retailer, increased its number of new or reactivated customers of 150% in 2014.
	Better customer segmentation, account management and customer insights	Retail; Utilities	20% to 25% increase of overall return for an electricity supplier in Europe through better customer relationship, lower cost to serve via digital channels, direct marketing and improved behavioural demand-response.
	Increased sales through better customer engagement	Retail; Utilities; Banking	
	Increased value of consumer loans	Banking	1.1% to 6.6% average increase in consumer loans for a European bank through Big Data and Analytics technologies.
Cost savings	Considerable cost savings because of more efficient stock management practices	Retail	Tesco, a British retailer, achieved yearly savings of £6 million through efficiency in stock management (i.e. reducing discounts in cases of over stocking)
	Cost savings because of more efficient stock management	Retail	
	Lower costs of policy intervention thanks to analysis of past data	Government	The British National Health Services obtained £200 million savings on one single category of drugs by reducing prescription of more expensive medicines.
	Reduction in product returns, scrap and rework	Manufacturing	Grundfos, a Danish pump manufacturer, obtained 20% to 30% reduction of product return in 5 years.
	Cost reductions thanks to single data portal and repository and fewer interfaces	Banking; Utilities	Estimated savings of €1.5 per year for European Banks through the introduction of a single data portal and repository.
Enhanced Operational Efficiency	Prevented returns of articles /Decrease return rates of semi-finished and finished products	Retail; Manufacturing	Otto Versand, a German retailer, prevented returns of approximately 2 million articles in 2013 thanks to BDA technologies.

Impacts	Benefits	Industry / Sector	Relevant Examples
	Reduced time and resources needed to use, clean, analyse and validate data	Government	Nottingham City Council reduced time for handling information from 3 months to 3 hours by matching information sources automatically and not by hand.
	Reduction in total equipment downtime	Manufacturing; Utilities	Volvo, a Swedish car manufacturer, increased the time of uninterrupted equipment functioning from 22 to 38 months.
	Production increase thanks to predictive maintenance programmes	Manufacturing; Utilities	
	Reduced bottleneck-cycle	Manufacturing	Syngenta's U.K. branch (agro-industry) reduced bottleneck cycle time in worth of £2.5 million.
	Reduction in time spent for trouble-shooting	Banking	Unicredit, an Italian bank, achieved a 70% reduction of time spent for trouble-shooting in three years.
	Reduction in time for running management reports	Banking	
Increased effectiveness	Increased effectiveness of policy measures (reduction in crime rates; Reduced traffic on public roads); more focused policy intervention thanks to better data analytics	Government	Reduced road traffic by nearly 25% (and subsequent increase of train traffic of train passenger traffic) per day after one year of combined use of GPS data and toll charges (Stockholm local government).
	Reduction in order delivery time	Manufacturing	Overall reduction in order delivery time from 3.5 to 0.8 days for the average European manufacturer by improving sales and operational planning through BDA technologies.
	Reduction in response time	Manufacturing	
	Reduction in time to market	Manufacturing; Utilities	Reduction in time to market (144 to 87 days) and new product launches' failures (12.5% to 9.7%) for the average European manufacturer by improving sales and operational planning through BDA technologies.
	Reduction in new product launches failures	Manufacturing	
Promoting ICT adoption and deployment, entrepreneurship and new ventures	Bridging the gap between ICT-Advanced and ICT Traditional Sectors	Agriculture (Precision Agriculture)	Several, ICT-laggard, Portuguese SMEs now using advanced farm management software from Agrivi – a Croatian agri-tech start-up. Positive impacts: increase crop yields up to 50% and 23% thanks to smart water management and improved plant variety selection respectively.
	Promoting new ventures and business collaboration	Agriculture (Precision Agriculture)	Salt & Lemon, an Italian SMEs initially active in the media industry, now successfully active in precision farming solutions with clear gains for its users in terms of cost savings

146

SMART 2013/0063 European Data Market

Impacts	Benefits	Industry / Sector	Relevant Examples
			(25% reduction of use of fertilizers), productivity (5% improvements thanks to crop increase, reduced environmental contamination, time savings (drones covering one hectare in 10 minutes versus 90 minutes normally taken by traditional farm machines)

Source: European Data Market Monitoring Tool, 2016

SMART 2013/0063 European Data Market

9.3 Key Findings

The rapid emergence of a European data market is steadily supported by a new wave of data-driven innovation, which in turn leads to deeper and more widespread adoption of data-related technologies, pioneered by Big Data and advanced analytics solutions and applications. While not new per se, these technologies are helping wide sectors of the European economy to make novel use of an ever increasing amount of data and, in the process, are becoming a key factor leading a growing number of European organisations toward higher levels of efficiency and competitiveness.

An initial attempt to categorise data-related benefits suggests rearranging these benefits along four main categories: revenues, costs, operational efficiency, and organisational effectiveness.

- Data-driven technologies are helping European organisations to **increase revenues**. This is particularly evident in customer-facing activities where the use of Big Data and analytics is proving to be fundamental in finding new customers or reactivating existing ones (as in the retail industry), or in better understanding their needs and requirements (utilities, banking, retail industries) and generating new turnover (banking industry). As an example, Morrison's, a British retailer, increased its number of new or reactivated customers of 150% in 2014 and, all in all, 20% to 25% increase of overall return for an electricity supplier in Europe through better customer relationship, lower cost to serve via digital channels, direct marketing and improved behavioural demand-response.
- Big Data and analytics can also drive significant **cost optimization**. The retail industry is capitalizing on new, more efficient stock management practices offered by a more sophisticated interpretation of existing data, while manufacturers benefit from lower levels of scrap and rework. We found, for example, that Tesco, a British retailer, achieved yearly savings of £6 million through efficiency in stock management (i.e. reducing discounts in cases of over stocking) thanks to the adoption of Big Data and Analytics technologies.
- **Operational efficiency** is attained in banking through a considerable reduction in the time spent on trouble-shooting, administrative practices, and risk-containment activities (such as fraud detection). Government agencies can look at severe budget cuts more comfortably as data-driven technologies enable them to execute complicated data cleaning and data validating operations more swiftly and with fewer resources. To quote another example, Nottingham City Council reduced time for handling information from 3 months to 3 hours by matching information sources automatically and not by hand.
- Big Data and analytics can **increase effectiveness** in utilities and manufacturing and improve key performance indicators such as order-delivery time, response time, and time to market, while the public sector can devote its (often scarce) resources more effectively toward prioritised policy needs thanks to better use of data analytics. For example, our research demonstrated that the manufacturing sector could obtain an overall reduction in order delivery time from 3.5 to 0.8 days for the average European manufacturer by improving sales and operational planning through the adoption of last generation Big Data and Analytics technologies.
- Data-driven technologies also allow traditional, non-ICT driven sectors to significantly upgrade their uptake and deployment of advanced technologies with positive impacts on all the above mentioned categories of benefits. As an

example, Big Data and Analytics can be applied to the agricultural sector giving birth to what is now called "Smart Agriculture" or "Smart Farming". Our research showed that SMEs active in the agricultural sector and other small farmers with a traditional low usage of ICT, can reap up to 25% savings in costs due to a reduced use of fertilizers, increase productivity of at least 5% thanks to crop increase, or, again, augment crop yields through smart water management and smart fertilization practices.

10. MEASURING THE DATA WORKERS SKILLS GAP

10.1 Definition

The **Data Worker Skills Gap** captures the potential gap between demand and supply of data worker skills in Europe. Demand and supply are estimated separately taking into account the supply of graduates, the level of unemployment, and the entry and exit flows of the data worker market.

The main goal of this indicator is therefore to verify the potential existence of a lack of supply of data skills in Europe which may become a barrier to the development of the data industry and the rapid adoption of data-driven innovation.

More specifically, we use the following definitions:

- **The supply of data workers** is equal to the data skills supply stock which **includes individuals employed as data workers, plus unemployed data workers.** The source of data skills is the education and training system and the move from other careers to the data market. Consistently with the data workers' definition, data skills include job positions that would often sit outside the remit of the IT department and be closer to the business-side of the organisation.
- The demand for data workers is the sum of existing and open positions, i.e. the demand for data workers includes the data workers employed plus the unfilled vacancies.
- The data skills gap occurs when the demand of data workers is higher than the supply (excess demand); when the opposite occurs there is over-supply of data skills.

As shown in the table below, 2016 saw in Europe an imbalance between demand and supply of 420,000 unfilled data workers positions, corresponding to 6.2% of total demand – up 0.3% from the same value for the year 2015[21]. In calculating the data skills gap in Europe for the year 2016 we have slightly adjusted the amount of the gap for the year 2015, which is now estimated at 393,000 positions overall (a minor fine-tuning of 3,000 units, downwards). Irrespective of this slight alteration, our new estimate for 2016 confirms that the data skills gap in Europe is on the increase, although at a slower pace than in the previous years - the gap almost doubled in 2015 vis-à-vis 2014 and it now grows by approximately 6% in the period 2015-2016.

By 2020, based on the updated demand and supply trends presented in this report, we foresee a continuing imbalance between demand and supply, under all 3 forecast scenarios. In the Baseline scenario the data skills gap will grow at something more than 16% over the next four years totalling a number of unfilled positions of almost 770,000

---

[21] The measurement of this indicator is based on a model built on several assumptions, particularly concerning the share of graduates who choose to become data workers and the entry and exit flows to the data workers labour market. The results should be considered as estimates. The results for the total EU28 are more reliable than the results by Member State. Official statistics from Member States are not available to clarify the specific dynamics of supply and demand by Member State and particularly the potential mobility of workers within Europe. The capture of this information by Member States should be encouraged for future analysis of the skills situation within the European Union.

in the EU28; the High-Growth scenario will see the gap expand considerably at a 60.7% pace with almost 2.8 million positions unfilled; conversely, under the Challenge scenario, the gap will actually diminish with respect to the year 2016 to reach 226,000 units, hence marking a negative CAGR of -14.4%. In fact, according to the Challenge scenario, in 2020 the data worker career will be less attractive for professionals engaged in other careers (restraining one of the main inflows of data workers supply) but also the demand for data workers will be negatively affected, due to a general cooling down of the overall economy. As a result, both the supply as well as the demand of data workers will suffer, with the latter diminishing at a faster pace than the supply. As a consequence, there will still be an excess of demand but at a lower level than in the Baseline or in the High-Growth scenario and the overall gap will be reduced to 226,000 units. This Scenario will present a very varied and divided picture where some of the largest and most advanced EU economies will still experience a considerable gap (this is the case of the UK, Germany and, to a lesser extent France) whereas weaker economies such as Italy and Spain will actually witness an oversupply of data workers. Fuelled by an intensified demand of data and by higher penetration rates of data-related technology, the High Growth scenario is characterized by a fast increase of the data workers demand which risks to lead to almost 2.8 million unfilled positions, equivalent to a share of 27% of total demand. In the Baseline scenario, the picture would be similar to what we presented in the previous release of this report although the overall gap will be on the increase, similar to what we estimated for the year 2016. According to this scenario, in fact, in 2020 the overall gap in the EU28 will represent 9.8% of the total demand of data workers – up more than 3% points with respect to our previous estimates.

Table 31 Indicator 5 Data Workers skills gap in the EU, 2014-2015-2016, 2015-2020, 000s and %, three scenarios

Indicator 5 — Description										
N.	Name	Description		Actual		Baseline Scenario		Challenge Scenario		High Scenario Growth
				2015	2016	2020	16-20 CAGR	2020	16-20 CAGR	2020	16-20 CAGR
5.1	Data Workers skills gap	Gap between demand and supply of data workers, N, 000s	EU27	341	348	541	11.7%	194	-13.6%	2,067	56.1%
			EU28	393	420	769	16.3%	226	-14.4%	2,797	60.7%
5.2		Gap between demand and supply of data workers, %	EU27	6.3%	6.4%	8.8%		3.7%		25.4%	
			EU28	5.9%	6.2%	9.8%		3.5%		26.8%	

Source: European Data Market Monitoring Tool, IDC, October 2016

10.2 Measurement Approach

The measurement of this indicator is based on a model combining the separate estimates and forecasts of the demand and supply of data workers with data skills. The definition and estimate of data workers is explained in Chapter 4. The total demand of data workers for the years 2014, 2015 and 2016 includes the data workers currently employed (as calculated by the Indicator 1) plus estimated vacancies per year (currently unfilled positions). The forecast demand of data workers to 2020 under the 3 scenarios presented in Chapter 4 is the total potential demand (it incorporates future potential vacancies). The supply is estimated aggregating the data workers currently employed, the unemployment rates, and the inflows and outflows to the data worker market (including retirements, changed careers, upskilling and so on). More details are presented in the Methodology Annex.

10.2.1 Updating the Measurement of the Data Workers' Skills Gap in 2016

The methodology approach is the same developed in 2014, 2015 and 2016 to estimate the supply-demand balance of ICT skills in the EU (e-Skill) 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 Brussels[22]. 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 5.1 and 5.2 the study team has adjusted the model combining the estimates and forecasts of the demand and supply of data workers with data skills leveraging a wealth of different sources, among which:

- OECD Digital Economy Papers, among which: OECD (2014), Measuring the Digital Economy: A New Perspective; OECD Publishing.
- ILOSTAT (International Labour Organization) Statistics and Databases (2015)
- EUROSTAT Tertiary Education Statistics (Last update: December 2015).
- European Data Science Academy (EDSA) project deliverables and publications (July 2015).

---

[22] "e-Skills in Europe: Trends and Forecasts for the European ICT Professional and Digital Leadership Labour Markets (2015-2020)", empirica Working Paper (November 2015)

Figure 46 The Data Skills Demand-Supply Balance Model
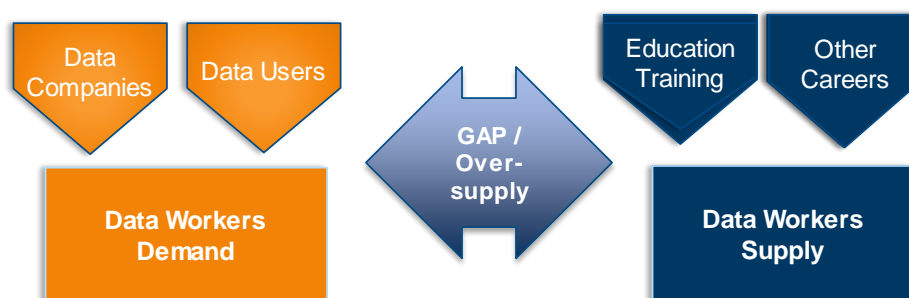
Source: European Data Market Monitoring Tool, IDC 2016

10.3 The Indicator by Member State

In 2016 the data workers' skills gap, as shown in the table below, is expected to increase in absolute terms under the Baseline scenario from 420,000 to 769,000 positions and from 6% to 13.8% as a share of potential demand. Considering the current rates of unemployment, this is not a huge gap even if it points to a structural demand-supply imbalance. This is confirmed by the additional increase of the gap under the other scenarios, particularly the High Growth where instead the size of the potential gap of more than 2 million unfilled positions would represent a real constraint for the development of the data market. Table 32 below summarizes the extent of the data workers' skill gap in units and share of potential demand for the largest Member States the total of the EU28.

Table 32 Data Worker skills gap by MS, 2016-2020, 000s, three scenarios

		FR	DE	IT	PL	ES	UK	Total Gap
2015	N. 000s	34	76	36	18	34	51	393
	% Gap	4.2%	6.0%	6.3%	4.2%	8.6%	4.1%	5.9%
2016	N. 000s	40	59	53	75	35	72	420
	% Gap	5%	5%	9%	15%	9%	6%	6%
Baseline Scenario	2020 N. 000s	23	60	67	73	2	227	769
	16-20 CAGR	-13%	0%	6%	-1%	-54%	33%	16%
	% Gap	3%	5%	10%	13%	0%	13%	9.8%

153

SMART 2013/0063 European Data Market

		FR	DE	IT	PL	ES	UK	Total Gap
Challenge Scenario	2020 N. 000s	37	95	-5	62	-7	32	226
	16-20 CAGR	-2%	13%		-5%		-14%	-14%
	% Gap	4.7%	7.7%	-1.1%	12.9%	-1.9%	15.5%	8.3%
High Growth Scenario	2020 N. 000s	324	309	411	198	180	730	2797
	16-20 CAGR	68.4%	51.2%	66.9%	27.6%	50.5%	78.6%	60.7%
	% Gap	26.6%	18.8%	40.4%	26.7%	28.7%	31.8%	26.8%

Source: European Data Market Monitoring Tool, IDC October 2016

The supply of data workers is less elastic than the demand, responding less quickly to the evolution of economic conditions in the 3 different scenarios, as shown by the table above. When the higher education system adds new skills, for example, the new graduates will entry the labour market with an average time lag of 5 years. However, the market responds to increased demand through additional training, learning on the job and inflows from other careers, and this is likely to counteract the potential gap. Internal mobility between the MS is also likely to be a factor, but the actual size of these internal EU inflows for specific categories of professionals is very difficult to estimate. On the other hand, we see that the growth of total demand has strong variations between the 3 scenarios, which require anticipated action by policy makers to counteract the potential data skills gap.

Figure 47 Total EU Supply-Demand of Data Workers, 2016-2020, 000s, three scenarios
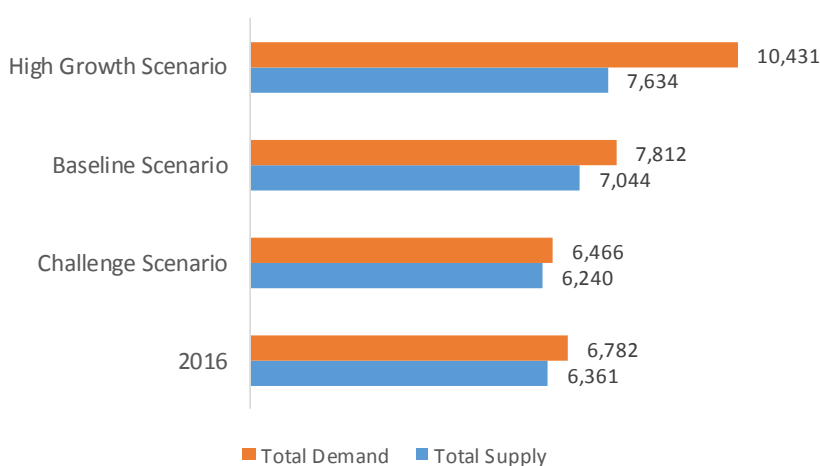
Note: demand includes potential vacancies

Source: European Data Market Monitoring Tool, IDC October 2016

The charts below present the data skills gap estimates by scenario for the major countries, highlighting the different dynamics of the data market as well as of the national labour markets. More specifically:

- The UK, Germany and France (leading data markets) show a mid-size gap in the Baseline scenario, because the positive dynamics of supply do not keep up with the strong growth of demand. In the Challenge scenario they still have a similar gap, because the increase of supply and demand from 2016 to 2020 is slower than in the Baseline scenario, but with a similar dynamic: so the gap found in 2016 remains also in 2020, although it will be more evident in Germany and less pronounced in the UK and in France. In the High Growth scenario, these Member States will all register the highest gaps across the EU because their market increases much faster than supply of data skills.
- Poland shows a data skills gap in line with those exhibited by most of the EU Member States across the three scenarios under consideration. The low growth of demand in the Challenge scenario results in a smaller gap than in 2016, while the opposite happens in the High growth scenario where the potential gap reaches a substantial level of total demand.
- Italy and Spain are characterised by high unemployment and dysfunctional labour markets. As a result, they show a moderate growth of demand in the Baseline scenario and Challenge scenarios, compared to a supply growth on a par with the average EU growth, resulting in small gaps in the former scenario and small oversupplies in the latter scenarios. On the contrary, both countries will suffer from significant data skill gaps should the High-Growth scenario become reality.
- Other countries, too would suffer from data skills gap in the High Growth scenario. However, in Spain the growth trend of students and graduates in Science, Technologies, Engineering and Mathematics (STEM graduates) is higher than in Italy so the Italian gap is much higher than the Spanish gap in the High Growth scenario.

Figure 48 Data Worker skills gap by Member State, 2020, Baseline scenario, 000s (total supply + unfilled positions = total demand)
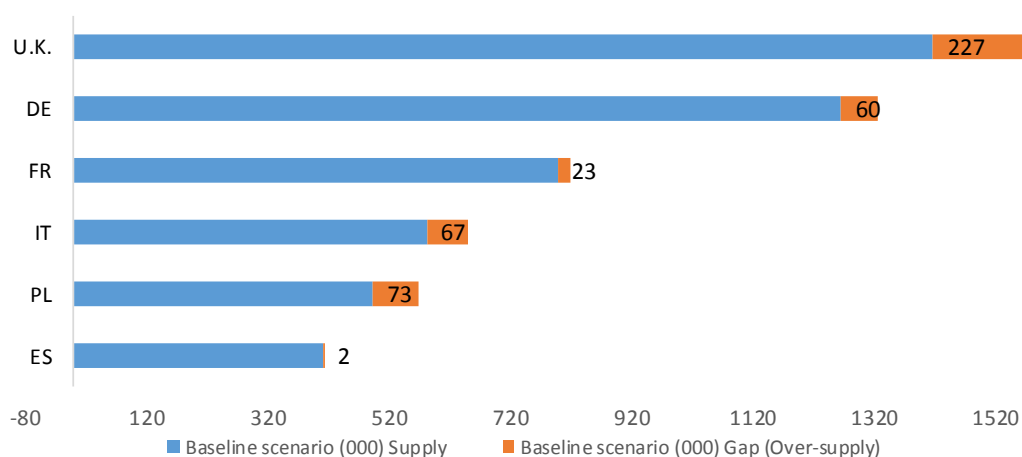
Note: demand includes potential vacancies

Source: European Data Market Monitoring Tool, IDC October 2016

Figure 49 Data Worker skills gap by Member State, 2020, Challenge scenario, 000s (total supply + unfilled positions = total demand)
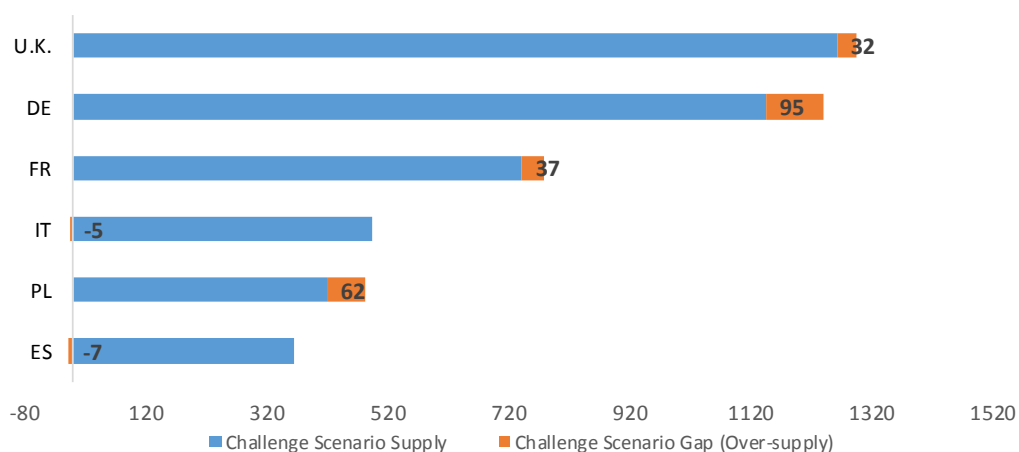
Source: European Data Market Monitoring Tool, IDC, October 2016

Figure 50 Data Worker skills gap by Member State, 2020, High Growth scenario, number (000s) (total supply + unfilled positions = total demand)
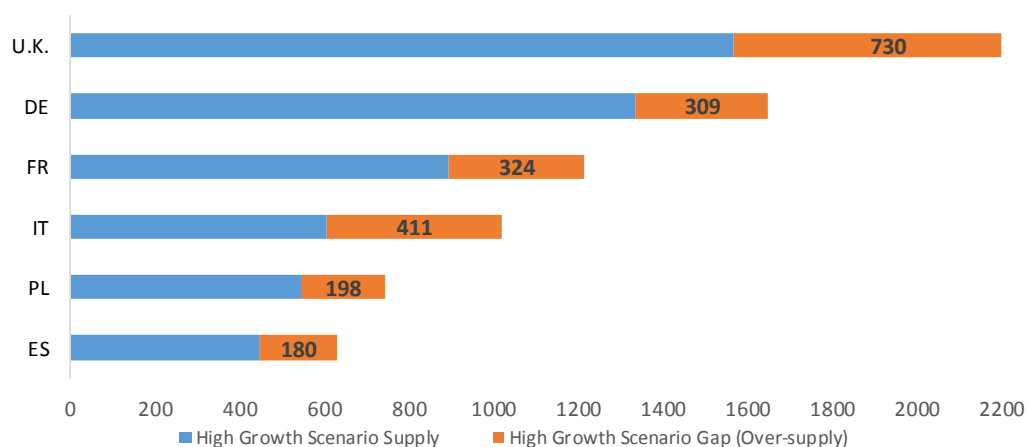
Source: European Data Market Monitoring Tool, IDC October 2016

10.4 Additional Insight on Big Data Analysts

Our definition of "data workers" is more comprehensive than the one in use by the OECD, which is limited to two professional categories and does not capture the full range of skills needed for data-driven innovation[23]. The EDSA project[24] definition of data scientist

---

23 Please see also the Methodology Annex, 15.2.5: Data Workers Definition 2016

24 The European Data Science Academy (EDSA) project, http://edsa-project.eu/

SMART 2013/0063 European Data Market

is closer to our "data worker" definition but more oriented to scientific and highly sophisticated skills, while less focused on business needs. For this reason, and to avoid any possible confusion, we will not use the term data scientist in this report, accepting EDSA's definition for it.

In this study we have adopted a broad definition of the data market, which includes but is not limited to Big Data technologies. Similarly, we have given a broad definition of data workers which includes but is not limited to Big Data analysts. There is still a lot of discussion on the precise profile and role of data scientists. What is certain is that they are expected to have sophisticated technical skills which have been in rare supply until now. Therefore, there is concern in the ICT industry that the rapidly increasing demand of data scientists may remain unsatisfied and this may become a bottleneck for the development of the data industry. For example, "statistical analysis and data mining" came in first in the top list of the 25 hottest job skills of 2014 compiled by LinkedIn on the basis of literally hundreds of millions of jobs requests worldwide[25]. In this paragraph we analyse the available evidence about the potential level of demand of these professional figures and provide an estimate at the EU level of the potential gap in 2020, within the context of the Baseline scenario[26].

Up to now, there are no consistent existing statistics to estimate the demand and supply for big data analysts, i.e. the people with deep analytical skills. However, several sources have applied different approaches to arrive at the demand side of the equation.

For example, according to the Forfas report (2014)[27], U.S. Bureau of Labour estimated that deep analytical talent in 2008 represented about 0.1% of overall employment in the U.S. By 2013, this proportion had increased to 0.21%. If these proportions would be applied to Europe, then this would mean that in 2014, there would be around 460,000 deep analytical skills jobs in the EU, or the equivalent of around 7% of data workers. However, the U.S. market is typically more advanced in use of technology than the European region as a whole. Typically, there is a delay in adoption rates of around 3-5 years. In addition, the take up of advanced analytical skills are more likely for larger companies than in small medium enterprises, which characterise a large proportion of the make-up of EU employers. Consequently, it would seem that these proportions would be too high to apply to Europe.

The study published by SAS and the Tech Partnership for Europe in October 2014[28], arrives at an estimate of demand for big data jobs in 2013 of 21,400 big data jobs. The report also estimated that by 2017, the demand would be for 47,600 jobs, resulting in an average annual growth rate of 22.1%. Applying this growth rate would give an estimate of around 26,000 jobs in the U.K. in 2014 – or around 2.2% of data workers. If we apply this proportion to total number of data workers in EU28, we arrive at 145,000 jobs in

---

[25] http://blog.linkedin.com/2014/12/17/the-25-hottest-skills-that-got-people-hired-in-2014/

[26] Existing statistics and other evidence on data analysts and data scientists is scarce and often inconsistent. As a result, we believe it is reasonable to limit our analysis to the Baseline scenario at this stage.

[27] Assessing the Demand for Big Data and Analytics Skills, 2013-2020, Forfas, April 2014

[28] Big Data Analytics Assessment of Demand for Labour and Skills 2013–2020, SAS and Tech Partnership, October 2014

2014 – a very different number from the estimate using the U.S. Bureau of Labour figures.

However, just as the U.S. is an earlier technology adopter than Europe overall, the U.K. is among the early adopter countries within the EU. According to IDC[29], spending on Big Data technologies and services in the U.K. accounts for 21.9% of the Western European Big Data market. If we take this as an indication also of the jobs related to Big Data – bearing in mind that Western European total IT spending accounts for around 90% of total IT spending in the EU (so therefore assuming this is a reasonable approximation), then we could arrive at an estimated number of Big Data jobs in the EU28 of 119,000 in 2014 or 1.8% of data workers.

The SAS and Tech Partnership report also estimates that the number of U.K. Big Data jobs in 2020 will reach 56,000. Compared to our 2014 estimates above, this would mean a compound annual growth rate from 2014-2020 of 13.5% in the U.K. However, it should be assumed that since the U.K. also is an early adoption country, the rest of the EU28 will grow faster over the period. If we assume that the U.K. only represents 21% in 2020, then we would arrive at an estimated 267,000 big data jobs demanded or the equivalent of 3.1% of data workers at a 14.3% CAGR over the period.

Is this a reasonable estimate? According to the IDC report sourced above, the expected CAGR for technology and services spending on Big Data from 2014 to 2018 is 22%. Similarly, one of the areas where deep analytical skills will play a big role – namely the Internet of Things – is reported to grow by high double digit growth rates by many sources out to 2020 and beyond. So, the estimate may be conservative – it is certainly unlikely to be wildly optimistic.

What of the supply side? Again, in this area there are even fewer statistics or estimations available. It is quite clear from analysing the entry requirement for relevant data science, big data or other deep analytical skills courses and degrees that a pre-requisite is typically significant computing, mathematics, engineering, science or statistics bachelor level degrees.

However, not all of the talent entering the big data profession will come directly from education system but rather have some years' experience in another field – but most likely still with the same degree background. If we consider the weight of the science, computing, mathematics engineering or statistics graduates on the supply of data workers in a given year, this varies from 1.1% in 2014 to 1.4% in 2020. However, as stated, there will also be people already in the job market that will retrain to gain the necessary qualifications and there is a huge emphasis especially in the IT industry on retraining employees for Big Data work. One could assume that most of the relevant graduates would be attracted to big data careers (since these also often carry a premium in remuneration). So an estimate of 1.5% of the supply in 2014 and 2.5% (to take into account the ramp up of the many new courses and graduate programmes on offer) in 2020 may be a reasonable assumption. Under these assumptions, we find the demand and supply for Big Data skills in the EU28 as shown Table 33.

---

[29] Western Europe Big Data Technology and Services 2011–2013 Market Size and 2014–2018 Forecast by Country and Segment, August 2014

Table 33 Potential Demand and supply for Big Data skills in the EU28, 2014-2020, 000s

EU28 – Estimated number of Big Data Analysts (000s)			
	2014	2020	CAGR 2020/2015 Baseline scenario
Demand	119	267	14.3%
Supply	94	201	13.6%
Forecast gap	26	66	16.9%

Source: IDC 2015

In summary, based on this analysis, we conclude that:

- Big Data analysts are a highly qualified, small group of professionals who (according to our preliminary estimates) in 2014 represented less than 2% of the estimated population of data workers in the EU, counting about 119,000 workers;
- The forecast demand of Big Data analysts to 2020 is expected to grow much faster than the demand for data workers, at a CAGR of 14.3%; supply is not expected to grow as fast, and therefore we find a potential supply-demand gap of 66,000 unfilled positions, corresponding to approximately 17% of demand.
- This forecast gap at 17% is almost three times as high as the data workers' gap estimated for the Baseline scenario in 2020 (approximately 6%), even if the absolute number for Big Data analysts is tiny compared to the millions of data workers.
- This appears to confirm the concern among Big Data stakeholders about the potential lack of Big Data skills and the need for Europe to catch up in the training and education for these skills.

These estimates were developed very quickly to respond to specific concerns on Big Data analysts' skills. Further research is needed to validate these estimates and provide more in-depth analysis.

10.5 Key Findings

This indicator measures the potential gap between demand and supply of data workers' skills in Europe, which are estimated separately taking into account the supply of graduates in the disciplines relevant for data workers, the level of current and structural unemployment, the entry and exit flows to the data workers market, and forecast demand for data workers' skills. The model requires a high number of assumptions and the results are more reliable for the total EU28 than for the individual Member States, because we lack sufficient evidence about the actual/potential mobility of data workers and markets demand-supply mismatches at national level.

The main findings are the following:

- According to our model, in 2016 there was a gap between total demand and supply of data workers of 420,000 unfilled data worker positions in the EU, corresponding to 6.2% of total demand. The year before, we estimated a total gap of 393,000 workers (approximately 5.9% of total demand), showing that the positive dynamic of the data market and growth of the data economy do bear an

159

influence on the availability of data skills in Europe. However, considering that the level of "natural" unemployment (which cannot be eliminated) is estimated at around 5%, the difference between data workers' demand and supply in both 2015 and 2016 cannot be regarded as a very large gap.

- The 3 scenarios highlight the diverse potential trajectories of the demand-supply balance of data skills to 2020. The structural imbalance between demand and supply resulting in a data skills gap appears in all 3 scenarios to 2020, but is projected to be a relevant problem only in the High Growth scenario where it could reach more than 2 million of unfilled data workers positions.
- The gap indicator varies substantially by Member State in the forecast scenarios because the growth rates of both demand and supply can be very different. Therefore, the gap indicator is very sensitive to specific national dynamics.
- In the Baseline and the Challenge scenario the gap is forecasted to be mid-sized, corresponding to 769,000 unfilled positions in the first or 226,000 in the second by 2020. However, in both scenarios the gap is concentrated in the large countries (UK, Germany and France) while Italy and Spain could be characterized by a limited over-supply in the Challenge scenario. This reflects the different conditions of the national labour markets with higher unemployment in Southern Europe and a less dynamic growth of the data market.
- Based on our definition, data workers comprise a wide portfolio of skills and may come from a wide range of disciplines, from Science, Technology, Engineering and Mathematics (STEM) to social science, business and law. There is a high number of graduates with the right skills who may step in if the data workers' career grows in attractiveness because of strong demand. The dynamics of supply of relevant graduates are projected to be positive across Europe. Apparently there should not be a supply problem of data skills for the user industries.
- The level of the gap appears to be very much influenced by demand-supply mismatches by country and industry, which enterprises will try to solve by training on the job and creating multidisciplinary teams. Therefore, this type of gap should not be a relevant constraint on the demand side, but is likely to be more relevant for the supply side (for the European data industry). There is in fact plenty of anecdotal evidence that enterprises have trouble sourcing highly specialised, technical data skills and this supply problem may remain relevant in the future.

11. MEASURING CITIZENS' RELIANCE ON THE DATA MARKET

11.1 Definition

This indicator was conceived to measure the level of citizens' reliance on data and to provide a more complete picture of the importance and social benefits of the data economy to the EU. It aims to complement the "business orientation" of the other indicators by providing a snapshot of how citizens are taking advantage of data-driven solutions in their daily life. The following paragraphs describe in detail the scope and a suggested measurement approach of the indicator, based on desk research, as well as an assessment of the availability of relevant data. This is a very innovative indicator which will be further improved over the next rounds of measurement of the European Data Market Monitoring tool.

11.1.1   Background: What to Measure

One of the key policy targets included in the data value chain strategy is to "increase citizens' use of data for informed behavioural decisions".

Data-driven solutions are present in many aspects of most citizens' lives: listening to music via Spotify, connecting to friends via social networks, commuting using mobile apps for information on public transport or traffic or even simply shopping at a local supermarket with a fidelity card. In this respect, the majority of citizens are already using services built on big data, but the technology remains mostly transparent and invisible to the user.

The challenge is to identify aspects of citizens' daily life that can be recognised as "data-driven," where the data gathering and analytics dimension have an impact on the decisions taken by citizens. We therefore exclude from the analysis activities for which there are consolidated indicators but that do not entail a significant "data-driven" dimension and decision-making function, such as:

- Information gathering activities, such as seeking health information online, which are not necessarily data-intensive;
- Transaction-focused activities, such as online banking which still for the most part is about performing payments rather than in-depth expenditure analytics;
- Services which use big data solutions to provide added value without showing it to users.

The goal of this indicator, therefore, is to measure to what extent citizens are taking data-driven decisions in their daily life.

In recent years, a trend has emerged concerning the use of computing devices to track one's own activities (from sleeping, to doing sport, from dieting to commuting) defined as "quantified self." The pervasiveness of ICT and sensors opens up new possibilities for tracking our actions and gaining information that was not previously available. How much sleep we take, how much exercise we do, how much electricity we consume at home, how much time we spend commuting, how much pollution we generate, and how government is spending our money is just some of the data that is now easily available to citizens.

The availability of this data triggers the behavioural change needed to address some of the key challenges of our time: healthier lifestyles, more sustainable transport choices, and more transparent governments.

Within these selection criteria we can already identify several application domains that are suitable to be measured:

- Health and wellness (or fitness) for tracking personal activity;
- "Fintech" apps to monitor and analyse personal spending patterns;
- Home automation solutions that help monitor energy expenditures or security alarms;
- Personal sustainability applications that allow for the tracking of personal environmental impact (e.g. by travelling);
- Analytics tools for open government data.

Ideally, our indicator should cover all these areas. Unfortunately, they are quite diverse and we have not been able to muster sufficient and comparable data thus far. We have to deal with the traditional dilemma between importance of the phenomenon in terms of actual and potential impact, and data availability.

Fintech apps are still in an emerging phase and seem to get increasing attention by market and investors pointing to future growth; however, official statistics online cover the usage of online banking which does not capture the essential novelty of data-driven solutions. Even within non-official statistics, it is not possible to identify comparable figures across EU Member States.

Home automation apps and sustainability apps are growing but are still a niche market; with very little statistics available.

In the case of open government data, the issue is certainly of primary policy importance as the EC and Member States have made substantial investments in transparency and open data portals. However, there are no official statistics on the percentage of citizens that accessed and used open government data and related analytics services. As a proxy, we could use the number of visits to the national open data portals, assuming that they reflect the citizens' usage of data, but even these data are not gathered in a comparative and systematic way across all Member States.

As of today, it is clear that the health and wellness domain solutions are more widespread than the other application areas. The health and fitness applications area is also an example of how the use of wearable computing and apps leads to the use of data to change behaviour (which is part of the "quantified self" trend).

Prevention and healthier lifestyles are widely recognised as the key determinants of health, in particular in relation to the growth in deaths caused by heart disease and obesity: six of the seven biggest risk factors for premature death – high blood pressure, excessive cholesterol, disproportionate Body Mass Index, inadequate fruit and vegetable intake, insufficient physical inactivity and alcohol abuse – depend on how people eat, drink and move. According to a recent study commissioned by DG SANCO, healthiest lifestyle would lead up to a three-year increase in life expectancy.

Today, there are plenty of ICT solutions (apps and wearable device) that aim to provide data on one's behaviour to track the amount of activity and sleep, as well as tracking biological data such as heartbeat. They also typically enable the user to share this data with friends (e.g., on Facebook). Knowing one's behaviour and being able to share and

162

compare with friends provides the motivation to actually change the behaviour and live healthier lifestyles. It is well recognised in behavioural studies that the community of peers has a major role in influencing one's behaviour. Health and fitness apps sector saw the strongest growth in both downloads and daily usage in 2014. In future, they are expected to increase growth as activity tracking is one of the main functionalities of the emerging smartwatch segment (such as the recently released Apple iWatch). The recent EC Green Paper on Mobile Health (COM(2014) 219) confirms that self-monitoring apps represent the most important segment of mHealth solutions and open up significant opportunities for the empowerment of patients and ensure quality of life through increased prevention, which is also linked to behavioural change. There is emerging evidence that those who track their health indicators also changed their approach to health.

11.1.2   Scope

In order to measure the adoption of health tracking, it is important to clearly define what is concretely meant by health tracking solutions. This is a complex trend involving many different potential services. To make it measurable and concrete, we identify mainly two product categories:

- Health and fitness apps on smartphones, which is the fastest growing apps category and includes apps that help track one's physical activity and sleep, such as Nike's FuelBand app which tracks your movement and converts it into a proprietary index of activity;
- Wearable computing that monitors physical activity, e.g., complex accessories such as Fitbit devices and smart accessories such as the Pebble smartwatch.

As the next step in the methodology, we have examined the availability of data on the diffusion and take-up of apps and wearable computing devices.

11.2 Indicator 6: Measurement

In order to measure the degree to which citizens make decisions based on data-driven solutions, we suggest using the two indicators identified above as follows:

- An indicator measuring the adoption, that is the % of citizens owning a wearable computing device, sourced from IDC;
- An indicator measuring the capability of citizens to use data in their daily life that is the average level of digital skills measured by the human capital index, sourced from the EC.

Our assumption is that the majority of wearable computing devices will be used for data-driven solutions and to make decisions based on them. The level of digital skills is a good proxy of citizens' capability to do so, and therefore to actually use data-driven solutions. We therefore assume that the combination of these two indicators will measure to what extent the population in each country is currently making decisions based on digital data.

We should also transparently state some of the limitations of such indicator:

- It does not capture the "health and fitness apps" sector. Unfortunately, no robust data are available for Europe on the number of people who installed such apps. However, we know that the vast majority of apps downloaded are for free or at very low cost (below 5 EUR), hence they do not reflect a strong interest in the

functionalities as the actual purchase of a wearable device (typically priced from EUR 50 to 250).

- It does not reflect that most likely the early adopters of wearables have more advanced skills than the average of the population. However, in the absence of robust data and evidence about the actual level of such skills we maintain the same proportion, also to reflect the difference between access and actual usage of the device, and to ensure future comparability of data.
- It does not include the additional "data-driven decisions" in other domains such as banking, government, transport, energy consumption. As such, it is biased in favour of countries with a population more interested in health issues and against countries with citizens which takes data-driven decisions in other domains.

These limitations suggest a probable underestimation of the final results, at least in the short term and until these devices reach into the mainstream.

The indicator is therefore produced for each country as:

Citizens data indicator for each country = Percentage of citizens owning a wearable device (calculated by dividing the number wearable devices sold in last 2 years by the number of citizens) x percentage of citizens with basic ICT skills.

11.2.1   E-Skills Indicator

Citizens' data skills are here captured using the indicator for basic or above basic ICT skills as measured by the Digital Economy and Society Index (DESI) as published in November 2016. This indicator is based on the Eurostat ICT Household survey. To be classified in this group, an individual has to possess basic or above basic skills in all the four Digital Competence domains included in the index: information, communication, content-creation and problem-solving. The indicator is shown in Figure 46 and reflects the distribution of digital skills across Member States. The EU average is 56%, similar to the one in in 2014 and 2015.

Figure 51 DESI Basic ICT skills Indicator by MS, 2016, normalised values

Source: https://ec.europa.eu/digital-agenda/en/desi

SMART 2013/0063 European Data Market
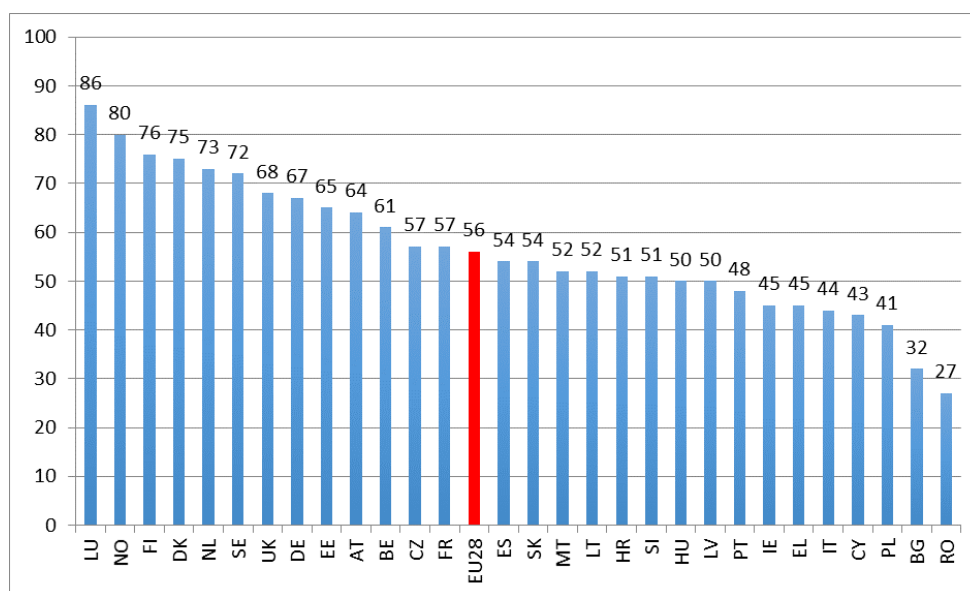
11.2.2 Ownership of Wearable Computing Devices

The following indicator shows the average number of citizens owning a wearable device per country. It is obtained by dividing the total market for wearable devices in the last two years (as calculated by IDC) by the number of citizens, considering an average replacement cycle of two years (similar to smartphones). IDC predicts the wearable market to grow strongly from $96 Million in 2013 to $15,817 Million in 2020. The data is available for 18 EU Member States, covering 95% of the EU population.

Figure 52 Percentage of citizen owning a wearable device, 18 Member States, 2016
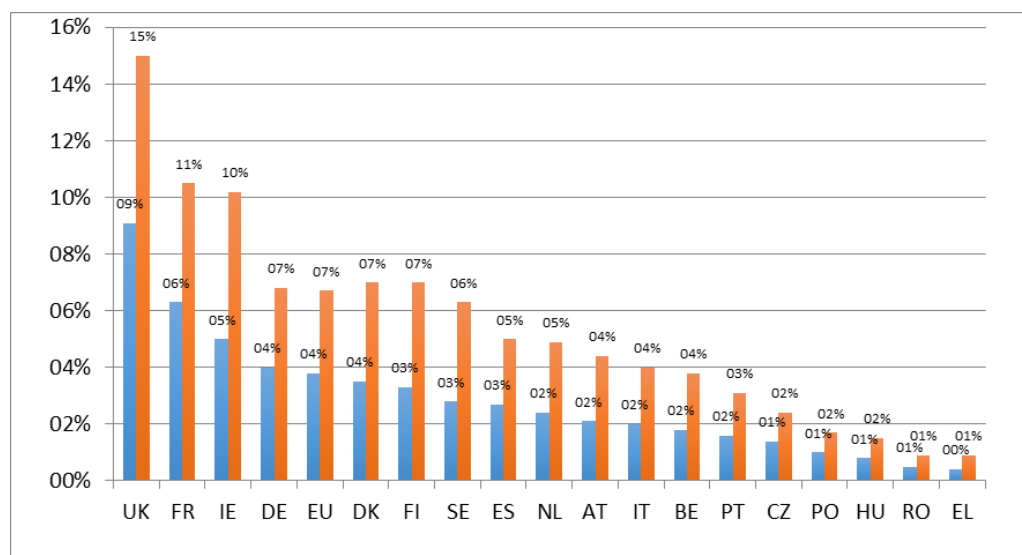
Source: Elaboration on IDC Wearable Computing Tracker, Blue is 2015, red 2016.
Notice that data of 2015 have been updated.

The data shows that the U.K. is the leader in terms of adoption of wearable computing devices, and its position strengthened in 2016. The largest economies appear to be more mature markets, with the exception of Italy. Smaller innovative countries such as Ireland and Finland show high uptake too. East European and Mediterranean countries tend to be laggards.

The average of these Member States is 6,7%: in other words, on average one on every fifteen citizens owns such device. This is the result of a strong growth across all countries: during the previous year the European average was 3,8%.

11.2.3 The citizens' data indicators

The summary indicator estimates the total percentage of citizens who take decisions and adapt their behaviour based on the data acquired through the wearable device. The values are obtained by multiplying the previous indicator by the percentage of people with basic ICT skills in the country.

Figure 53 Percentage of citizens taking decisions based on data, selected Member States, 2016
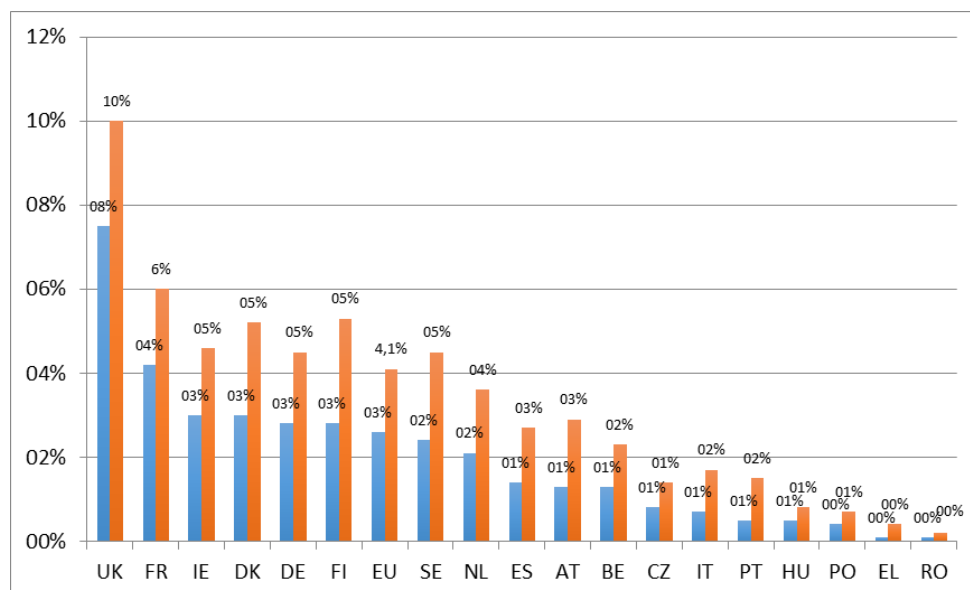
Source: Elaboration on IDC Wearable Computing Tracker, Blue is 2015, red 2016.

The overall average for EU18 is 4,1%, up from 2,6% in 2015 and 0,6% in 2014, indicating that a small, but fast growing minority of citizens today take behavioural decisions based on big data. The amount is growing continuously showing no signal of slowing down. The U.K. is a clear leader with 10%, almost doubling the percentage of the second MS (France). Eastern European and Mediterranean Member States in particular remain at the bottom of the ranking, as they have both relatively low rates of citizens with basic ICT skills *and* low adoption of wearable devices[30].

Table 34 Summary of key sub-indicators for Indicator 6, EU18

Indicator 6	2015	2016
% of population with basic ICT skills	58%	56%
% of population with wearable device	3.8%	6,7%
% of population taking decisions based on data	2,6%	4,1%

These data reflect a very early stage of the wearable market, but at the same time confirm the expected fast growth.

---

[30] Based on the Baseline scenario developed in this report, and assuming a similar rate of growth in basic ICT skills to the one experienced from 2009 and 2014, we expect a rapid increase of the percentage of citizens using data to take decisions in the EU14 over the next five years. Because of scarceness of data at this stage we did not estimate this indicator in 2020.

11.2.4 Targets: increase citizens' use of data for informed behavioural decisions

The availability and visibility of data our behaviour, gathered through unobtrusive devices such as apps and wearables, is expected to trigger more informed decisions on our lifestyle (health, travelling, nutrition etc.). In addition, data generated on a daily basis by wearable devices may play a role in addressing some of the key societal challenges highlighted by Horizon 2020, which should be addressed, in turn, to implement the policy priorities of the Europe 2020 strategy. In particular, the indicator on citizens' reliance on data should help addressing the following challenges:

- Health, demographic change and wellbeing;
- Europe in a changing world - inclusive, innovative and reflective societies.

Table 35 EDM Indicators and key policy targets (1)

Key Policy Target	EDM Indicator
Increase citizens' use of data for informed behavioural decisions **Address the following challenges:** - **Health, demographic change and wellbeing;** - **Europe in a changing world - inclusive, innovative and reflective societies.**	Indicator 6.1: % of citizens taking data-driven decisions in their lifestyle, calculated through the proxy of the percentage of ICT skilled population owning a wearable device.

Available data show a strong growth of wearables adoption in EU in 2016 vis-à-vis the previous year, growing from 3,8% to 6,7% of the EU population. Considering the IT literacy rates, this tells us that today about 4,1% of the EU population takes data-driven decision in their daily life choices. These are conservative estimates since:

- early adopters are more likely to be IT literate;
- we do not have data on health and lifestyle app adoption, which also are used for taking data-driven decisions.

This strong growth also validates the methodological choice done one year ago to focus on wearables as a proxy of data-driven decisions by citizens.

In this respect, the overall European landscape is changing fast: two years ago we were referring to an interesting but relatively marginal phenomenon; today wearables are becoming more and more popular among larger portions of EU citizens. Thanks to the increasing adoption of smart wristbands, smart watches and similar products, we can assume today and a wider array of decisions pertaining to healthy and active living, but also to other areas (daily commuting and transportation, dining, hospitality and entertainment, for example) is becoming more and more data-driven or, at least, is significantly affected by the availability and use of new, available data.

With a back of the napkin calculation, considering the expected growth rates for both wearables adoption and it literacy, we can assume that by 2019[31] 20% of the EU

---

[31] IDC sources on which the sub-indicator "Percentage of population with wearable device" is construed are available until the year 2019, hence the decision to build our reasoning within this timeframe.

population will take data-driven decisions in their daily life (25% of citizens' adoption and 80% literacy rates). Since lifestyles affect major societal aspects such as health and productivity, as well as economic aspects such as public finances, this adoption level can be anticipated to have a major visible impact on the EU economy and society as a whole.

Still, it is clear that data do not forecast a rapid, exponential growth to reach the majority of the population. As things stand, the growth is likely to continue with a similar pace in the future. Adoption rates are very different among countries. While the UK stands out for its high adoption rates (10% of the population) in Romania only 0,2% of the population take data-driven decision. Geography remains in fact a key factor to explain adoption. Central and Nordic countries perform better and occupy the top 8 positions. Mediterranean countries and Eastern European countries occupy the places at the bottom of the ranking.

Worryingly, there is no sign of convergence, perhaps because of the early stage we are at. Countries with higher adoption in 2015 grew faster, as it is dramatically shown by this scatter plot.

The expected rapid growth in the adoption of wearable devices, if confirmed, is likely to generate important cascading effects as public and private providers will have to adapt their services and products to a more lifestyle-aware demand, and to even more important overall societal changes: 6 of the 7 biggest risk factors for premature death – blood pressure, cholesterol, Body Mass Index, inadequate fruit and vegetable intake, physical inactivity and alcohol abuse – depend on how people eat, drink and move and recent studies estimate that the full adoption of healthier lifestyle could lead to an increase of 3 years in life expectancy.

Nevertheless, taking full advantage of these opportunities implies addressing several of the challenges identified in the Digital Single Market: it's important to ensure that Europeans have adequate ICT skills to use and make sense of data (4.3.1 of the DSM strategy), and to avoid lock-in and promote the interoperability of data and devices (4.2), especially in key areas such as health.

11.3 Policy implications

From the point of view of public policy, the data relate to citizens' data-driven decisions are particularly meaningful. 2015 is the year when we had the confirmation that data-driven decisions are becoming an important aspect of daily life. This is supported by the positive trend in the following year. In the UK, already today about 10% of the population does so. If confirmed, the fact that citizens increasingly make up their mind in taking decisions based on data is likely to affect all aspects of human life, and all vertical policies, from health to consumer protection to environmental protection.

On the one hand, this opens up significant opportunities for public policy. Wearable devices and in general health and fitness apps provide both information and psychological "nudges" that increase the control of citizens over their health and actually change their behaviour towards healthier lifestyles, with the deriving benefits for the society and for the public finances. For instance, there is robust scientific evidence that "the use of a pedometer is associated with significant increases in physical activity and

significant decreases in body mass index and blood pressure[32] and that "a simple smartphone app significantly increased physical activity over 8 weeks in a primary care population."[33]

This could extend to other domains and help address some of the most important challenges we face. For instance, improvements in measuring the environmental impact of our daily choices could help the adoption of more environmental friendly choices, hence helping to address one of the most critical issues faced by humanity. The wide adoption of wearables would also radically enhance the possibility of research and innovation. First, it will provide unprecedented amount of data on human behaviour, that could help investigating and understanding it. Secondly, it will enable a new generation of wearable apps being developed with entirely new solutions that cannot even be imagined today. In other words, it will open new markets.

On the other hand, some policy issues deal not with the opportunities, but with the bottlenecks. It is clear that IT and data literacy are necessary to take full advantage of these data. Citizens will be empowered only once they have the necessary skills to critically analyse and interpret the data; otherwise they will remain passive users of the information provided by apps and devices. The opportunities of data driven decisions at this stage remain open only to a small minority of the population, and could worsen already existing social divides for instance with regard to healthy lifestyles. This is true not only at social but also at geographical level. The greater the opportunities, the greater the risks of increasing geographical disparities in grasping them. The gap between leading and lagging countries is growing, and should be closely monitored for the future. Last but not least, data protection and usage issues remains crucial. The immense opportunities related to the growth of personal data availability means that citizens should have greater control over the data gathered by these devices, about who can access and use them.

11.4 Key Findings

This analysis shows that there is a wide range of possible indicators to be used for measuring the "data maturity" of citizens; in most cases there is a major data gap to adequately support decisions. However, in some cases data could be made available quite easily (such as the access data to open data portal by citizens). There are several market segments of data-driven consumer solutions, ranging from health and wellness, to financial services, home automation, security and open government data. However, the health and wellness area is the most relevant one in terms of uptake, impact on

---

[32] Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA. 2007;298(19):2296–304.

Harris T, Kerry SM, Victor CR, Ekelund U, Woodcock A, Iliffe S, et al. A Primary Care Nurse-Delivered Walking Intervention in Older Adults: PACE (Pedometer Accelerometer Consultation Evaluation) -Lift Cluster Randomised Controlled Trial. PLOS Med. 2015;12(2):e1001783 doi: 10.1371/journal.pmed.1001783

[33] Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. The British journal of general practice. 2014;64(624):e384–91. doi: 10.3399/bjgp14X680461

behaviour and policy relevance. Health tracking concretely refers to health apps and wearables that enable analysis of one's activities, and enable citizens to take data-driven decisions on their lifestyle.

In the absence of data regarding to what extent citizens take decisions based on data, we formulate an indicator based on two key aspects that determine data-driven decisions based on skills and access to the technology. Based on our feasibility analysis we have identified two basic indicators measuring these aspects, potentially available for all Member States and periodically updated:

1. For adoption, that is the % of citizens owning a wearable device, estimated based on the number of devices sold (sourced from IDC);
2. For skills, the % of citizens with basic ICT skills as measured by Eurostat and described in DESI.

The underlying hypothesis is that Member States with a high diffusion of health tracking devices and high level of skills will have a high share of citizens making decisions based on data-driven solutions.

The results show that in 2016 only 4,1 % of EU18 population used data provided by wearables to drive their decisions, varying from 10% in U.K. to 0.2% in Romania.

While these data are relatively low, this is a very new phenomenon and the expected growth is strong. Based on IDC market prediction, we estimate that by 2019 more than 20% of EU citizens use data to take decision on a wide array of daily activities such as doing sport, dieting, commuting, holiday-making and entertainment, just to mention a few.

The U.K. is at the forefront of adoption of this technology. Eastern European and Mediterranean Member States are at the bottom of the ranking. Even more worrying is that the countries with higher adoption in 2015 are those who grew faster, hence widening the gap.

As per 2016, these data are likely to marginally underestimate the phenomenon as the skills of the early adopters are expected to be more advanced and the indicators do not capture the diffusion of health apps, for which reliable data are not available.

12 WORLDWIDE MONITORING OF THE DATA MARKET

12.1 Introduction

The data economy is in a truly global phenomenon — not only are investments in Big Data, business analytics software, and other data-related technologies rapidly expanding outside the traditional ICT-oriented regions of Europe and North America, but data itself is a production factor than can be easily transferred from region to region and from country to country. As a result, monitoring the data market in Europe is not sufficient and, in this study, we have chosen to extend our analysis to three main EU partners:

- **The United States**: representing a high-income economy[34] in the North American region as well as the EU's top trading partner[35] and the country accounting for approximately a third of overall ICT spending worldwide[36].
- **Brazil**: an upper-middle-income economy representing one of the most successful emerging markets in the world as well as one of the countries with the highest ICT Development Index (IDI) in Latin America[37].
- **Japan:** the largest high-income economy in the Asia/Pacific region, as well as the main Asian EU trading partner after China and a mature ICT market with a number of similarities with the EU as a whole.

For the Final Study Report, we have leveraged the same secondary sources that we used to obtain the results in the first round of measurements. We have accessed the latest statistical sources available and complemented our investigation with existing IDC research and ad-hoc secondary research on data-related technologies in the three countries under consideration. Nevertheless, due to the wide disparity of the available statistical sources for these EU partners, we have kept our international focus on a restricted set of core indicators, in particular[38]:

- Indicator 1.1: number of data workers;
- Indicator 1.2: employment share of data workers;
- Indicator 2.1 number of data companies;
- Indicator 3.1: revenues of data companies;
- Indicator 4.1: value of the data market;
- Indicator 4.2: value of the data economy;
- Indicator 4.3: incidence of the data economy on GDP.

---

[34] The World Bank, Countries and Lending Groups, http://data.worldbank.org/about/country-and-lending-groups#MENA

[35] Client and Supplier countries of the EU28 in Merchandise Trade (2015), Directorate General for Trade, European Commission, Last update, April 2016; http://trade.ec.europa.eu/doclib/docs/2006/september/tradoc_122530.pdf

[36] IDC Worldwide Black Book Pivot Table, Q1 2016 Update, IDC, April 2016

[37] "ITU, Measuring the Information Society Report 2015" (PDF). Geneva, Switzerland: International Telecommunication Union (ITU), 2015

[38] In this study, we do not present Indicator 6.1: Citizens' Reliance on Data Market for the United States, Brazil and Japan because of the lack of sufficient and/or consistent evidence of the two fundamental components of this indicator, i.e.: e-skills indicator and the ownership of wearable computing devices.

SMART 2013/0063 European Data Market

The first part of this chapter presents each of the above indicators for the United States, Brazil, and Japan separately, while in the second part an overview of the above indicators across all three international partners and a comparison with the EU28 will be provided.

12.2 Approach to Estimates of the International Indicators

The current estimate of the main indicators for the three select EU international partners leverage IDC databases available at worldwide level. Data such as ICT spending is available for most countries worldwide and is gathered with the same approach across the board – these data series are perfectly comparable at international level.

Other issues, however, continued to emerge when updating the indicators for the chosen EU international partners. For example:

- Unavailability of some specific data series used for the EU;
- Lack of information and data gathered with the survey, which was conducted only at European level.

12.2.1   Updating the International Indicators for the year 2016

To obtain the updated indicators for Brazil, Japan and the U.S., the study team conducted ad-hoc desk research on a variety on both internal and external sources, including but not limited to:

- IDC's worldwide and regional detailed market forecast estimates for IT Hardware, Software, and IT Services from 2014 through 2016;
- IDC Worldwide Black Book (Standard Edition), quarterly updates form the years 2014 through 2016. The Black Book represents IDC's quarterly analysis of the status and projected growth of the worldwide ICT industry in 54 countries.
- IMF World Economic Outlook (WEO) Database, October 2016
- Consensus Forecasts, Consensus Economics, monthly updates, September 2015 – October 2016.

12.2.2   Number of Data Workers

To estimate the number of data workers outside the EU we could not count on the ISCO data series by ISCO groups (1 digit), for the three countries. In fact, the most reliable data we had in our overall model for Brazil, Japan, and the U.S. was the data market estimates based on the IDC data sets. As a consequence, we have estimated the data workers using the data market, equivalent to the data revenues, as a starting point. The approach was similar to the one we used to estimate the data workers in the forecast estimates for the EU.

The key assumptions that we applied for the three EU partners are:

- There is one global "technology," i.e. a global production function for the data products and services.
- This means that in such a production function, labour and technology are substitutes. Thus, each country may employ a different proportion of labour and capital to achieve the same level of data market.

- The ratio between the TFP (Total Factor Productivity) of data and TFP of the general economy varies for each country according to its capital investments in ICT and the efficiency of the country.

The data workers are estimated as a production factor which depends both on the ICT investments achieved in the country in recent years and on the TFP of the country and the data industry.

12.2.3   Size of the Data Market

The size of the data market is estimated based on IDC's Black Book data, which gives detailed historical and forecast data for the worldwide ITC market, coupled with IDC's hardware and software worldwide tracker data for the countries estimated: Brazil, Japan, and the U.S. This is the same process used to size the data market for the Member States. The data market is based on a contribution from business analytics software, system infrastructure software, IT hardware (servers, storage, and networking hardware), and IT services associated with the data market. Selected segments of business analytics software were used as not all components of BA software are used in the data market. Selected shares and tie ratios established from previously published IDC research on the Big Data market completed estimates for the three international countries.

More specifically we included data on:

- Business analytics, system management, and various software applications.
- Hardware associated with the data market including server, storage, and some networking.
- IT services ranging from training, education, and planning to maintenance, operations, and support. The estimate was based on a model calculating a varying share of the value of these market segments and then aggregating them to calculate the total market value. The various shares were calculated based on a combination of IDC's Big Data research, the surveys carried out for this study, and analyst expertise. The most recent version of the IDC Black Book database used for this model was released in September 2016 and provides historical data for 2013–2014-2015, estimated data for 2016, and forecast data for 2017–2020.

12.2.4   Value of the Data Economy

The estimate of the data economy in the EU was based on some of the components in the results of the survey. The survey conducted in the EU provided some quantitative information about the benefits to enterprises when using data products and services. Such quantitative information was not available for the three non-EU countries.

Consequently, it is at this stage impossible to estimate the forward indirect impacts, which are also the most relevant in terms of quantitative impacts.

As explained for the EU estimates, our approach sub-divides the economic impacts in different components. For Brazil, Japan, and the U.S. we are in a position to estimate:

- The direct impacts
- The backward indirect impacts

The other impacts would require additional field investigations.

12.3 The Data Market in the U.S.

12.3.1   Overview

America's ICT industry has been suffering after decades of intensified globalization, offshoring of production capacity, and exploitation of low-cost labour in emerging economies, but remains alive and vibrant. The U.S may no longer be the absolute leader of the world's ICT ecosystem. However, it is still home to the likes of Apple, Google, Oracle, Dell, IBM, Microsoft, VMWare, Amazon, Facebook, and Intel, to name just a few, and still accounted for more than a quarter of total spending on ICT in 2014, or almost a third if we exclude the telecommunication sector.

This dominant position is clearly reflected by the size and depth of the data economy in the U.S. According to the McKinsey Global Institute, in 2009 all sectors in the U.S. economy had at least an average of 200 terabytes of stored data for each company with more than 1,000 employees[39]. Again, in terms of technology investment spending, IDC estimates that total spending for business analytics software in the U.S. in 2014 accounted for nearly 45% of the total business analytics market worldwide. This is expected to increase to about 55% if we consider the more specific market for Big Data technology and services in the same year[40]. According to the new Worldwide Semi-annual Big Data and Analytics Spending Guide from IDC[41], worldwide revenues for big data and business analytics will grow from nearly $122 billion in 2015 to more than $187 billion in 2019, an increase of more than 50% over the five-year forecast period. The new Spending Guide expands on IDC's previous forecasts by offering greater revenue detail by technology, industry, and geography. From a geographic perspective, more than half of all big data and business analytics revenues will come from the United States. By 2019, IDC forecasts that the U.S. market for big data and business analytics solutions will reach more than $98 billion. The second largest geographic region will be Western Europe, followed by Asia/Pacific (excluding Japan) and Latin America.

The strength of the U.S. data economy is confirmed by the vibrant Big Data–related and business-analytics-related start-up scene. The amount of venture capital funding invested in U.S.-based software vendors addressing Big Data and analytics requirements grew from $155 million in 2009 to $877–$893 through August 2013, according to IDC; this is accompanied by significant funding at federal level for high-performing computing (HPC) and other data-driven technologies, which are proving to be a fundamental asset in securing the U.S.'s leading position in an ever globalised world.

A snapshot of the key data market indicators in the U.S. is outlined in Table 36 below.

---

[39] Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, McKinsey & Company, 2011

[40] Worldwide Big Data Technology and Services, 2014–2018 Forecast, IDC, September 2014

[41] Worldwide Semiannual Big Data and Analytics Spending Guide, May 2015

Table 36 Summary of Indicators: United States of America

U.S.A – Indicators' Overview

N.	Name	Metrics	2013	2014	2015	2016	Growth rate 2014/ 2013	Growth rate 2015/ 2014	Growth rate 2016/2015
1.1	Number of Data Workers	Total Number of Data Workers (000)	9,931	10,457	11,636	12,732	5.3%	11.3%	9.4%
1.2	Data Workers employment share	Share of data workers on total employment (%)	6.9%	7.1%	7.8%	8.4%	3.6%	9.4%	7.6%
2.1	Number of Data Companies	Total number of data companies	273,535	277,821	283,340	289,556	1.6%	2.0%	2.2%
3.1	Revenues of Data Companies	Total revenues generated by companies specialised in the supply of data-related products and services (M€)	€ 97,237	€ 103,935	€ 115,519	€ 129,173	6.9%	11.1%	11.8%
4.1	Value of the Data Market	Estimate of the overall a value of the data market (M€)	€ 97,237	€ 103,935	€ 115,519	€ 129,173	6.9%	11.1%	11.8%
4.2	Value of the Data Economy (Only Direct and Backward Indirect impacts)	Direct Impacts (M€)	€ 93,708	€ 99,398	€ 109,246	€ 108,521	12.2%	9.9%	-0.7%
		Backward Direct Impacts (M€)	€ 3,528	€ 4,536	€ 6,273	€ 7,270	19.3%	38.3%	15.9%
4.3	Incidence of the Data Economy on GDP (Only direct and backward indirect impacts)	Ratio between value of the data economy and GDP (%)	0.6%	0.6%	0.7%	0.8%	3.4%	11.8%	9.3%

Source: European Data Market Monitoring Tool, IDC 2016

12.3.2   Measuring Data Workers

In 2016, we saw a 9.3% increase of the U.S. data workers on the previous year, corresponding to 12.7 million that more than double the EU estimate (6.1 million) for the same year. The estimate is in constant increase from the year 2013, but at slower pace from 2015 to 2016 than in previous periods. As in Europe, these workers are involved in the production or use of data as a main part of their activity, as we are not counting full-time data workers. In terms of penetration, we estimate it to be approximately 6%, more than twice as much as in the EU28. This higher penetration in the U.S. is due to the fact that in the U.S. the diffusion of data is at a more advanced stage when compared with the EU; this is supported by an industry structure where large enterprises are more widespread than in the EU. As explained in the chapter about data workers, it may be difficult for SMEs to have teams with the necessary data skills, while for the large enterprises this appears to be easier and this is probably a success factor in the adoption

SMART 2013/0063 European Data Market

process of data in the U.S. In terms of data workers' employment share (the percentage of data workers on total employment), the U.S. 2016 estimate is considerably higher than the previous years, well above 8%. This represents almost three times the EU values (3.1% in 2016 and 2015, 3.2% in 2014 and 3.0% in 2013) confirming the strength of the U.S. data market, as well as the overall structure of the U.S. economy where traditionally ICT-intensive and data workers-intensive sectors, such as ICT, financial services, and professional services play a fundamental role.

12.3.3   Measuring the Value of Data Companies' Revenues and Data Market

We estimate the data market (as defined in Chapter 4) and the value of the revenues generated by data companies in the U.S. to increase at a growth rate of 11.8% year-on-year in 2016. The U.S data market totals therefore slightly more than EUR 129 billion in 2016, versus EUR 115 billion in 2015, nearly EUR 104 billion in 2014, and almost EUR 98 billion in 2013 — meaning the U.S. data market is more than double the size of Europe's and it is projected to grow at a faster rate than in the EU28 in the period 2016-2015 (11.8% in the US in 2015-2014 versus 9.5.0% in the EU). The prominence of the U.S. in the ICT sector overall, and in the development and adoption of data-driven technologies, is clearly confirmed by the sheer size of its data market. This prominence is also reflected in total number of data companies that we estimated in the country. In 2013 they were more than 273,000; they exceeded 277,000 in 2014 reaching 283,000 in 2015. The estimate for 2016 indicates almost 290,000 data companies in the US, showing a continuous growing trend of approximately 2% over the years, and 14% more than in Europe.

The U.S. economy, though, is characterised by a company structure which is heavily skewed towards medium and large enterprises and it is perfectly plausible that the U.S. necessitates of a proportionally lower number of companies than the U.S. to sustain a similar level of data companies' revenue and data market.

12.3.4   Measuring the Data Economy

Based on our estimates, the impacts derived from the production and adoption of data products and services in the U.S. are higher than they are in the other economies, including Europe. Overall, direct and backward indirect impacts represent 0.8% of GDP in 2016, indicating a constant increase through the years, but slower in its growth in the period 2015-2016. However, our estimates show a slight decrease in the direct impacts from 2015 to 2016 (-0.7%). This relevance of the impacts depends both on a wider dimension of the data market and on a higher incidence on the GDP. As explained for the European data economy, the direct impacts are generated by the data industry itself; they represent the activity engendered by all businesses active in data production. Such impacts are about twice those of the European direct impacts, showing that data products and services are, not surprisingly, in a more advanced stage of production and diffusion in the economy.

The backward indirect impacts represent the business growth resulting from changes in sales from suppliers to the data industry. Such impacts are, again, more than three times higher than the backward indirect impacts in the EU.

12.4 The Data Market in Brazil

12.4.1   Overview

Brazil is currently going through a deep recession. The country's growth rate has decelerated steadily since the beginning of this decade, from an average annual growth of 4.5% between 2006 and 2010 to 2.1% between 2011 and 2014. The GDP contracted by 3.8% in 2015. The economic crisis - coupled with the political crisis now facing the country - has contributed to undermining the confidence of consumers and investors. The crisis was further exacerbated by commodity price drops and the deterioration of investor sentiment with regard to emerging markets[42]. According to IDC, Brazil's total ICT expenditure did increase in 2015 of approximately 10% year-on-year passing from $147 billion in 2014 to more than $153 billion in 2015. In local currency, however, the increase is much more modest, as the Real lost at least 8% of its value against the U.S. Dollar in 2014 and an its continuing its descent in 2015. As a result, investments for information and communication technologies also came to a halt and the outlook for the year 2016 is likewise adverse.

Notwithstanding the negative economic trends, coupled by the increased political uncertainty, the data economy continues to play an important role in the country, with big Data, business analytics, and other data-related technologies continuing to attract a substantial part of Brazil's ICT spending. There is already a significant Big Data and analytics market in Brazil, which is related primarily to the financial, retail, and telecommunication industries. Brazil also has vast energy resources and an immense amount of data is generated to support its oil production and extraction activities. Since 1997, the government has opened the sector to competition and has favoured national and foreign investments in this space. Much of Brazil's data-related revenues, however, are generated by companies from outside the country (infrastructure or basic application providers) or by large consulting firms that are implementing big ICT projects. Big Data and other data-driven technology markets are still very much in an early phase in Brazil and it is reasonable to expect that many local companies could take advantage of this in the years to come. The development of the data economy in Brazil is still hindered by several factors. Among these, two are particularly worth mentioning here: as in other regions, the country suffers from a lack of qualified professionals. To tackle this, many of the country's top universities have launched new executive or part-time courses focused on teaching data analysis and other data-related skills. Interestingly, much of the data-related training is offered to professionals who do not have an IT background, but rather a statistics or math background.

In addition, Brazil seems to experience the phenomenon of the so called "Big Data-Washers-proliferation" more intensively than other countries. Companies that are labelling their solutions "Big Data" regardless of whether they really deliver on that promise or not are in fact a common reality in Brazil. This trend will continue as this is typical of many emerging technologies. Nevertheless, Brazil appears to be steadily heading toward a more sophisticated data economy, although this will take than what previously expected.

---

[42] The World Bank, 2016; http://www.worldbank.org/en/country/brazil/overview

Table 37 Summary of Indicators: Brazil

Brazil – Indicators' Overview

N.	Name	Metrics	2013	2014	2015	2016	Growth rate 2014/ 2013	Growth rate 2015/ 2014	Growth rate 2016/ 2015
1.1	Number of Data Workers	Total Number of Data Workers (000)	1,006	1,031	1,026	1,160	2.5%	-0.5%	13.1%
1.2	Employment share	Share of data workers on total employment (%)	2.1%	2.1%	2.1%	1.8%	2.2%	-0.5%	-14.0%
2.1	Number of Data Companies	Total number of data companies	34,346	34,840	35,456	35,979	1.4%	1.8%	1.5%
3.1	Revenues of Data Companies	Total revenues generated by companies specialised in the supply of data-related products and services (Million EUR)	€ 4,703	€ 5,289	€ 5,272	€ 6,049	12.4%	-0.3%	14.7%
4.1	Value of the Data Market	Estimate of the overall a value of the data market (Million EUR)	€ 4,703	€ 5,289	€ 5,272	€ 6,049	12.4%	-0.3%	14.7%
4.2	Value of the Data Economy (Only Direct and Backward Indirect impacts)	Direct Impacts (Million EUR)	€ 4,521	€ 5,289	€ 5,477	€ 6,157	12.2%	3.6%	12.4%
		Backward Direct Impacts (Million EUR)	€ 182	€ 217	€ 293	€ 290	19.3%	34.8%	-1.0%
4.3	Incidence of the Data Economy on GDP (Only direct and backward indirect impacts)	Ratio between value of the data economy and GDP (%)	0.2%	0.2%	0.4%	0.2%	15.0%	56.5%	-55.2%

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market

12.4.2   Measuring Data Workers

In 2016 the number of data workers in Brazil, approximately 1.1 million, is higher than in previous years Thanks to more positive market conditions, this trend shows a 13% growth over 2015. Data workers represent almost 2% of total employment in the country – a lower share than the 3.10% EU average for 2016.

12.4.3   Measuring the Value of Data Companies' Revenues and Data Market

The value of data companies' revenues and the data market in Brazil totalled EUR 4.7 billion in 2013 EUR 5.3 Billion in 2014, EUR 5.2 Billion in 2015, and jumped to 6 billion in 2016 with a 14.7% increase from 2015 to 2016.

The estimate of the total number of data companies slightly increased over the years; while it was a little over 34,000 in 2013, it reached nearly 36,000 units in 2016. This still relatively low number is destined to growth significantly over the next few years given the high pace growth registered by the values of data companies' revenues and data market in the country and its fast growing ICT spending.

12.4.4   Measuring the Data Economy

The data market and revenues in Brazil are still moderate. During the past few years, Brazil has invested significantly in ICT, and data products and services are benefitting from this. Nevertheless, these impacts are slowing down when compared with the previous year due to the economic crisis that has severely hit the country. All in all, however, the overall impacts of the data economy they are still positive in Brazil, even if the 2016 estimates show a minor contraction (-55% on the previous year), decreasing to the values of 2013 and 2014.

12.5 The Data Market in Japan

12.5.1   Overview

Japan's ICT spending totalled $242 Billion in 2014 and $243 Billion in 2015 – a very moderate growth. The country, however, remains the largest ICT market in Asia/Pacific after China. In fact, the rapid development of ICT has underpinned Japan's economic growth for the last few decades and Japan's total R&D spend as a proportion of its GDP has remained at the top among other industrialised countries and Japanese companies have spent the most on ICT among high-income countries worldwide.

The Japanese ICT sector has its strengths in FTTH (fibre to the home), high-definition image technology, home network, and mobile equipment technology, where the country enjoys a clear competitive advantage thanks to its supremacy in optics and imaging, components for mobile telephones and TV, and advanced visual content. That said, Japan's ICT market has been experiencing modest growth over the past few years and is expected to do so also in the near future with flat growth rates over the period 2015–2019.[43] This, coupled with the contracting economy in 2013 and 2014, has somewhat

---

[43] *IDC Worldwide Black Book Pivot Table, Q1 2016, IDC, April 2016*

hindered the data economy in Japan. There are other reasons for this: firstly, the market for business analytics software is still at an early adoption stage in Japan; secondly, tight security and privacy policies do not favour widespread use, processing, and storage of data; and thirdly, data providers as a whole are still relatively few and far between in the country. Key players such as NTT-Data/Twitter Japan, Culture Convenience Club (retail), and Japan Railways do exist but their room of manoeuvre is limited by tight security policies, among other factors.

Yet there are signs of modest recovery ahead: over the next three years, IDC expects the number of companies using Big Data technologies to expand significantly from a small circle of innovators with deep tech skills to a larger number of early adopters, including a substantial amount of large enterprises. This change of pace is seen in a number of business sectors and activities, including the retail sector and its marketing campaigns, with digital marketing activities cantered on individual customers' experiences, that are rapidly gaining momentum in Japan, fuelled by the shift from feature phones to smartphones and other mobile devices. This is expected to encourage the development of Big Data analytics designed to allow a wider range of consultants, advertising agencies, and business units within companies to utilise self-service data-driven technologies. This, in turn, may help stimulate the domestic economy in the run-up to the 2020 Tokyo Olympic and Paralympic Games. The key data market indicators for Japan are displayed in Table 38.

Table 38 Summary of Indicators: Japan

Japan – Indicators' Overview

N.	Name	Metrics	2013	2014	2015	2016	Growth rate 2014/ 2013	Growth rate 2015/ 2014	Growth rate 2016/ 2015
1.1	Number of Data Workers	Total Number of Data Workers (Thousands)	3,353	3,344	3,613	3,740	-0.28%	8.1%	3.5%
1.2	Data Workers employment share	Share of data workers on total employment (%)	5.3%	5.3%	5.7%	5.8%	-0.90%	0.4%	2.7%
2.1	Number of Data Companies	Total number of data companies	95,705	95,919	99,001	101,612	0.2%	3.2%	2.6%
3.1	Revenues of Data Companies	Total revenues generated by companies specialised in the supply of data-related products and services (M€)	€23,021	€22,228	€24,013	€25,513	-3.4%	8.0%	6.2%
4.1	Value of the Data	Estimate of the overall a	€23,021	€22,228	€24,013	€25,513	-3.4%	8.0%	6.2%

SMART 2013/0063 European Data Market

N.	Name	Metrics	2013	2014	2015	2016	Growth rate 2014/ 2013	Growth rate 2015/ 2014	Growth rate 2016/ 2015
	Market	value of the data market (M€)							
4.2	Value of the Data Economy (Only Direct and Backward Indirect impacts)	Direct Impacts (M€)	€ 22,250	€21,367	€2,928		-4.6%	7.3%	19.5%
		Backward Direct Impacts (M€)	€ 771	€ 860	€ 1,085	€ 1,189	11.6%	26.1%	9.6%
4.3	Incidence of the Data Economy on GDP (Only direct and backward indirect impacts)	Ratio between value of the data economy and GDP (%)	0.5%	0.5%	0.7%	0.9%	2.1%	35.4%	43.4%

Source: European Data Market Monitoring Tool, IDC October 2016

12.5.2 Measuring Data Workers

In 2016, the number of data workers in Japan shows a 3.5% increase over the previous year exceeding 3.7 million, which is about half as many as in the EU28. The data workers in Japan represent a 5.8% share of the total employment.

The Japanese economy was as well characterized by a slowdown in 2015, both in the GDP and in the investments trends. Nevertheless, the unemployment is very low in this country and despite the negative trend of the GDP, the employment continued to increase. The data market increased significantly in 2015 (+8%) and this had a positive impact on the number of the data workers.

12.5.3 Measuring the Value of Data Companies' Revenues and Data Market

Japan's overall data market and value of data companies' revenues was estimated at over EUR 23 billion in 2013, slightly decreasing at EUR 22 billion in 2014, and increasing again at EUR 24 billion in 2015. For the year 2016, we have estimated the value of the data market and of data companies' revenues to be more than EUR 25 billion.

According to the International Monetary Fund,[44] real GDP declined in 2014 and 2015 in Japan and IDC estimated very moderate ICT growth in in the period 2013-2015. Both GDP and ICT spending forecasts, though, are expected to improve steadily in 2016 and beyond, indicating that the value of Japan's data market will continue to growth in the near future.

The total number of data companies is estimated at approximately 95,000 in 2013, 96,000 in 2014, 99,000 in 2015, and exceeding 101,000 units in 2016 – a moderate but still positive trend reflecting the mixed performance of the Japanese data economy over the past three years. Our estimates show that the number of data companies has grown at a slower pace in 2016 compared to the past year.

12.5.4   Measuring the Data Economy

In 2016 direct impacts and backward indirect impacts account for 0.9% of GDP in Japan, versus 0.7% in 2015, 0.5 in 2014 and 2013. The estimates indicate a 43% growth on the previous year, explained by higher investments in ICT compared to recent years.

12.6 Key Findings

According to our first and second round of monitoring, the European data market and economy in the period 2013–2016 was consistently second to the U.S. in value and, to a lesser extent, in growth. In terms of impact of the data market on GDP, however, in the same years both the U.S. and Japan had higher levels of incidence. In each of these countries the dynamics of the data market are positive and the growth outlooks are promising.

12.6.1   International Comparison of Data Workers

The total number of data workers in the U.S. and their share of total employment are almost twice as high as in the EU, with penetration rates close or above 7% throughout the period, versus EU penetration rates of approximately 3% during the same years. This is clearly a result of greater data-driven innovation in U.S. enterprises and a more mature data market.

The growth rate for the data worker population is four times higher in U.S. than in Europe in the period in 2016. The data worker population in Japan is also quite high at over 3.7 million in 2016 and the share of total employment is also higher than in Europe at 5.8%. This reflects the high level of IT adoption in Japanese businesses.

Brazil has shown a moderate increase of the total number of data workers in 2016, being now at 1.1 million, with a penetration rate on total employment at around 2% and a growth rate in the last two years considerably lower than in Europe.

---

[44] World Economic Outlook, International Monetary Fund, April 2016

Table 39 Data Workers, International Comparison, Number, Growth, and Share of total employment

	2014 (000)	2015 (000)	2016 (000)	Y-o-Y Growth 2014/2013	Y-o-Y Growth 2015/2014	Y-o-Y Growth 2016/2015	% on empl. 2014	% on empl. 2015	% on empl. 2016
EU	5,818	6,005	6,161	0.8%	3.2%	2.6%	3.0	3.1	3.1
USA	10,457	11,636	12,732	5.3%	11,3%	9.4%	7.1	7.8	7.6%
Brazil	1,031	1,026	1,160	2.5%	-0.5%	13.07%	2.1	2.1	-14%
Japan	3,344	3,613	3,740	-0.3%	8.1%	3.5%	5.3	5.6	2.7%

Source: European Data Market Monitoring Tool, IDC October 2016

Figure 54 Number of Data Workers, International Comparison, 2016, Units (000a)

Number of Data Workers, 000s
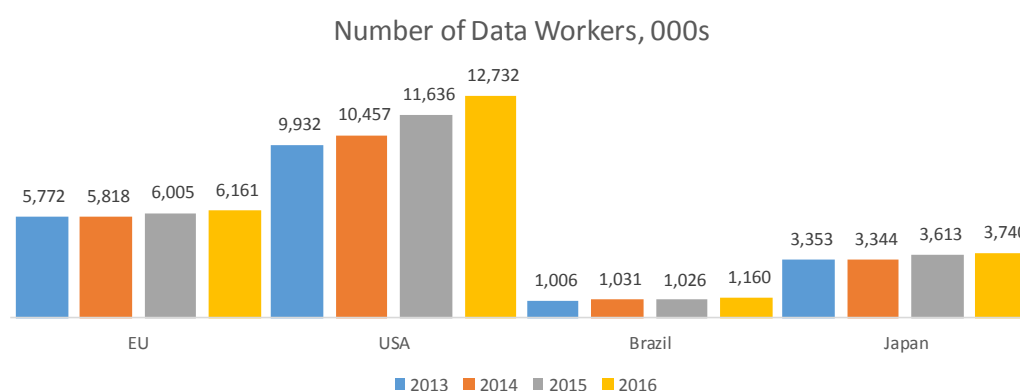
Source: European Data Market Monitoring Tool, IDC October 2016

12.6.2 International Comparison of Data Market Value

The overall value of the data market in the U.S. was estimated at EUR 97 billion in 2013, EUR 104 billion in 2014, EUR 115 billion in 2015, and EUR 129 billion in 2016 – twice as much the size of the EU data market and almost four times as much as Japan's over the three years under consideration.

A comparison of growth rates is also interesting, with the Brazil leading the way, well recovering from the previous negative year, followed by the U.S and the EU growing a little less.

Table 40 Data Market Value, International Comparison, 2013–2014-2015-2016, EUR Billion

	2014 EUR Bn	2015 EUR Bn	2016 EUR Bn	Y-o-Y Growth 2014/2013	Y-o-Y Growth 2015/2014	Y-o-Y Growth 2016/2015
EU	50	54	59	6.4%	7.0%	9.5%
USA	104	115	129	6.9%	11.1%	11.8%
Brazil	5.3	5.2	6.0	12.8%	-0.3%	14.7%
Japan	22	22.2	25.5	-3.4%	8.0%	6.2%

Source: European Data Market Monitoring Tool, IDC October 2016

Figure 55 Data Market Value, International Comparison, 2013–2016 EUR Billion
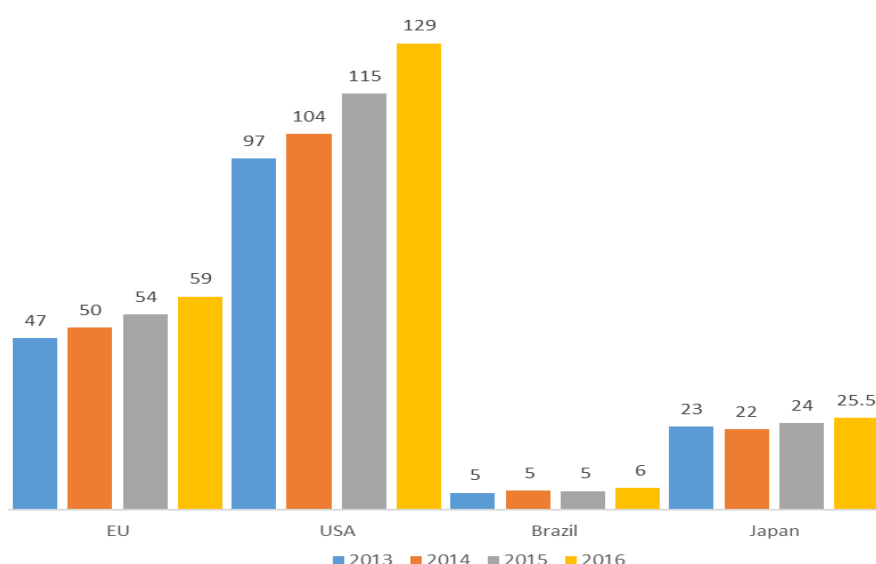
Source: European Data Market Monitoring Tool, IDC October 2016

A better relative assessment of the level of development of the data market is shown by the comparison of the share of the data market value on total ICT spending. The data market is one of the most dynamic components of ICT investments and a growing share is a good proxy of the fast adoption of innovation.

In the U.S. the relative weight of the data market on the overall ICT spending increased steadily from 2013 to 2015, passing from 9.8% to 12.3%; however, the estimate for 2016 indicates a minor decrease to 10.9%. In contrast, Japan exhibited a more varied picture reflecting the country's economic trend – the data market represented 9.2% of the overall ICT investments in 2014 (down 0.3% vis-à-vis the previous year) but it rebounded in 2015 achieving a noteworthy 11% according to our latest estimates; in

SMART 2013/0063 European Data Market

2016 our estimates show a very moderate contraction to 10.9%. Notwithstanding its difficult economic and political situation, Brazil managed to maintain a growing trend of its data market size in comparison to the overall country's ICT spending, which in 2016 reached 6.0% and almost doubled the percentage presented in 2015 (3.8%). With a data market amounting to a 9.5% share of the overall ICT investments in 2016, Europe still lags behind both the U.S. and Japan.

Figure 56 Data Market as a % of total ICT spending, International Comparison, 2013–2014-2015-2016, %
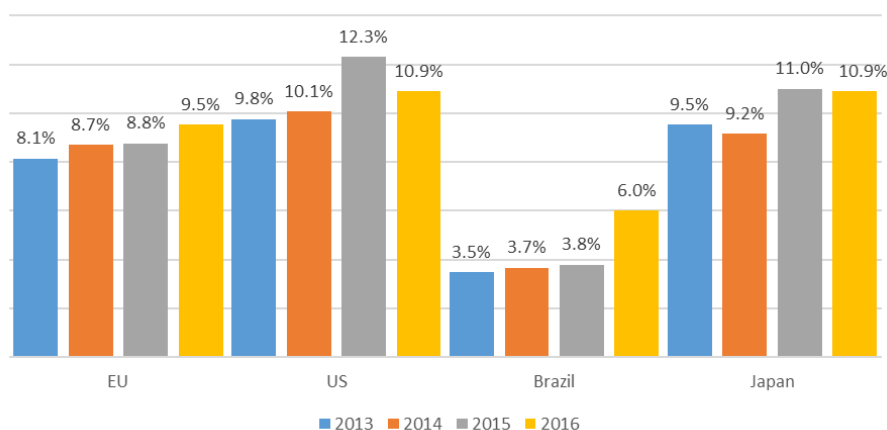
Source: European Data Market Monitoring Tool, IDC October 2016

12.6.3   International Comparison of the Data Economy as a Percentage of GDP

We have used the direct impacts (of the data market) and the backward indirect impacts (of the data industry) as a percentage of GDP in the period 2015-2016 to carry out a comparison of the data economy between the EU28 and the three countries taken into consideration in this chapter.

The results are quite interesting: in 2016 Japan took the leadership with a significant 0.9% of GDP generated by the data market in 2016, followed by 0.78% for the US. Europe still presents an incidence of economy of approximately half the one of the U.S. On the other hand, Brazil, which quite well over the past three years, shows a contraction in 2016. Brazil decreased the incidence of its data economy of 0.16 percentage points in 2016. All in all, the above landscape reflects the different levels of maturity of these economies, as well as their stage of data-driven innovation. What is more, the results of both this and the previous round of measurement of the international indicators, confirms the strong potential of the data economy, together with its ability produce tangible benefits for Europe if a rapid development of the data ecosystem is adequately supported.

Table 41 Data Economy as a % of GDP, International Comparison, 2013–2016, %

	2013 %	2014 %	2015 %	2016 %	Y-o-Y Growth 2015/2014	Y-o-Y Growth 2016/2015
EU	0.34%	0.35%	0.37%	0.37%	-5.6%	1.7%
USA	0.58%	0.60%	0.71%	0.78%	11.8%	9.3%
Brazil	0.20%	0.23%	0.36%	0.16%	56.50%	-55.24%
Japan	0.5%	0.5%	0.7%	0.9%	35.4%	43.4%

Source: European Data Market Monitoring Tool, IDC October 2016

Note: Data Economy here includes only the Direct and Backward Indirect Impacts

Figure 57 Data Economy (Total Impacts) as a % of GDP, International Comparison, 2013–2014-2015-2016, %

Source: European Data Market Monitoring Tool, IDC October 2016

SMART 2013/0063 European Data Market
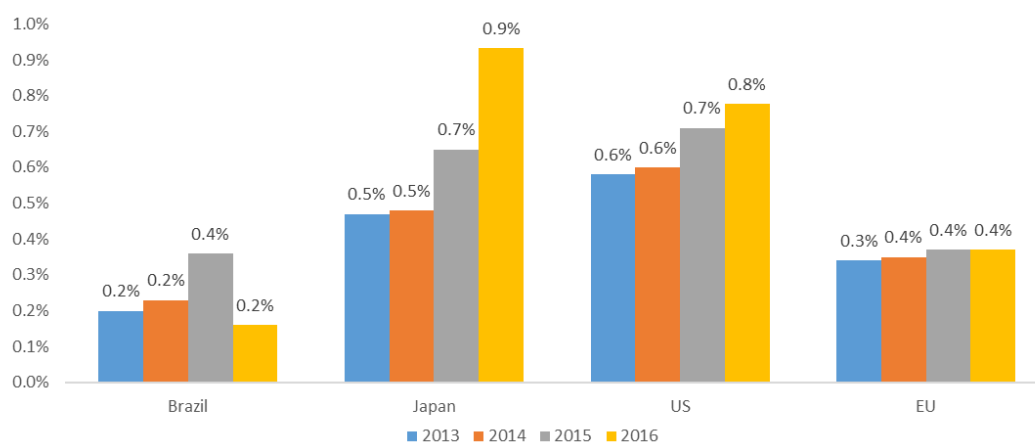
13 ASSESSMENT OF PROGRESS ON POLICY TARGETS

13.1 Main Policy Targets

The ultimate goal of this study is to define, assess, and measure the European data economy, supporting the achievement of the Data Value Chain policy, which aims to develop a vibrant and innovative data ecosystem of stakeholders to drive growth in the market in Europe. Since the start of this study the European Commission has further developed its Big Data and digital economy strategies. The main policy Challenges and the relative actions were identified in the EC communication "Towards a Thriving Data-Driven Economy" in July 2014 and were further enhanced by the launch of the European Digital Market strategy (DSM)[45] in May 2015, which aims at: 1) creating an inclusive single market for digital services and products to offer better access for consumers and businesses to digital goods and services across Europe; 2) creating the right conditions and a level playing field for digital networks and innovative services to flourish; 3) maximising the growth potential of the digital economy.

Within this context, the EC refined its approach towards a vibrant data-driven economy by launching a Communication package in April 2016 aimed at improving the digitisation process of the European Industry[46]. The digitisation of the European Industry would serve as coordination platform for other European, national & regional initiatives (such as Industrie 4.0 in Germany, Smart Industry in the Netherlands or the Nouvelle France Industrielle in France) and is accompanied by other initiatives in the field of standards[47] (to promote widely accepted standards in priority areas such as 5G, Cloud Computing, Internet of Things and Cybersecurity), accelerating the digitisation of public services[48] through a rejuvenated eGovernment Action Plan and reinforce the uptake and the strengthen the benefits to be derived by Cloud technologies[49] in a data-driven economy. At term, all these initiatives will produce results that need to be measured at regular intervals to gauge their progress towards their specific objectives and take corrective actions if necessary.

The implementation of a European Data Market Monitoring Tool responds to this need by determining the size and trends of the European data market while monitoring the interactions between the main stakeholders. This report presents the third set of full measurements and analysis of the European data market. Besides the obvious, which is the provision of indicators assessing progress toward measurable targets, the EDM Monitoring Tool will be able to feed information and intelligence into the policy process. More specifically, the usability of its results is indicated below.

---

[45] A Digital Single Market Strategy for Europe - COM(2015) 192 final

[46] Digitising the European Industry – COM(2016) 180 final

[47] ICT Standardisation Priorities for the Digital Single Market – COM(2016) 176 final

[48] EU eGovernment Action Plan 2016-2020, Accelerating the digital transformation of government – COM(2016) 179 final

[49] European Cloud Initiatives – COM(2016) 178 final

SMART 2013/0063 European Data Market

Table 42 Overview of the Data-driven Economy Communication Action Plan and EDM Monitoring Tool contribution

Community building	EDM Monitoring Tool contribution
1. A European Public-Private Partnership on Data	Input to design and measure KPIs
2. Digital entrepreneurship and open data incubator	Better understanding of digital entrepreneurship through the stories and the data industry analysis
3. Developing a skills base	Assessment of the demand for data skills and the demand-supply skills gap
5. Identification of sectorial priorities for R&I	Analysis of potential social and economic benefits of data-driven innovation

Developing framework conditions	
Availability of data and interoperability	
1. Fostering Open Data policies	Assessment of relevance in the development of the data market and data industry
2. Data handling tools and methods	
3. Supporting new open standards	
Enabling infrastructure for a data-driven economy	
1. Cloud computing	Assessment of relevance in the development of the data market and data industry
2.E-infrastructures and High Performance Computing	
3. Networks/ Broadband /5G	
5. Public Data Infrastructures	
Regulatory issues	
1. Personal data protection and consumer protection	Assessment of potential impact of regulatory barriers on the data market growth
2. Data-mining	
3. Security	
4. Ownership/transfer of data	

13.2 Data Policy Targets

The Data Value Chain unit of DG Connect has established a set of indicators to measure progress toward the achievement of its policies. We present here the first round of these measurements.

Table 43 EDM Indicators and key policy targets (1)

Key Policy Target	EDM Indicator
Number of European data companies, including startups, increased in the 2014-2020 period (2014-2020)	Indicator 2 — Number of Data Companies and forecast, by MS, by company size, by industry
Revenue generated by European data companies increased both in absolute and relative terms in the 2014-2020 period (2014-2020)	Indicator 3 — Revenues and Forecast of Data Companies, number and share of total revenues, by MS and by company size
Increase the number of data-related startups and fast-growing SMEs;	Anecdotal evidence — see EDM Stories and Data Landscape
Increase the revenue generated based on data in the Member States	Indicator 3 — Revenues and Forecast of Data Companies, number and share of total revenues, by MS and by company size

According to our latest round of measurement, the number of data companies in the EU28 is projected to grow between 5% (CAGR 2020-2015, Baseline scenario), 1% (Challenge scenario), and 8.9% (High Growth scenario). The Monitoring Tool includes the detailed forecasts under the three scenarios for EU28 and by company size, including the share of data companies that are SMEs (98.9% in number, 72.4% in revenues).

The revenues generated by data companies are expected to grow between 20.6% (CAGR, High Growth scenario), 10.4% (CAGR, Baseline scenario) and 4.7% (Challenge scenario) by 2020. The share of revenues is measured on the total revenues of all the companies in the ICT industry and the professional services industry, where we identified data companies. This share was at 3.4% in 2015, but was not extrapolated to 2016 and further to 2020, because of the lack of data on the denominator (global revenues of all the companies in the 2 industries where data companies can be found, ICT and professional services).

Concerning the number of start-ups and innovative SMEs, the survey of data companies did not yield sufficient evidence to calculate a separate indicator on this specific subject.

Table 44 Indicator 3 Total Revenues of Data Companies, EU, 2013 – 2014-2015-2016

Indicator 3 — Data Companies' Revenues								
N.	Name	Description	2013 €M	2014 €M	2015 €M	2016 €M	Growth Rate 2015/2014	Growth Rate 2016/2015
3.1	Total revenues of data companies in the EU	Total revenues of the Data Companies calculated by Indicator 2	47,727	51,656	56,033	61,819	8.5%	10.3%
3.2	Share of data companies' revenues	Ratio between data companies' revenues and total companies revenues in sectors J and M	3.0%	3.2%	3.4%	N.A	7.2%	7.4%

Source: European Data Market Monitoring Tool, IDC October 2016

Table 45 Forecast Indicators 2 and 3, Data Companies, Revenues and Data Users, 2020, three scenarios

	Forecast Indicators 2 and 3 – Data Companies, Revenues in the EU, 2020							
N.	Name	Actual		Baseline Scenario		Challenge Scenario		High Scenario	Growth
	Indicators	Value 2015	Value 2016	Value 2020	20-16 CAGR	Value 2020	20-16 CAGR	Value 2020	20-16 CAGR
2.1	Number of data companies (000)	249,100	254,850	310,250	5.0%	265,250	1.0%	359,050	8.9%
3.1	Total revenues of data companies in the EU (€ Mill)	56,029	61,819	91,874	10.4%	74,158	4.7%	130,708	20.6%

Source: European Data Market Monitoring Tool, IDC October 2016

13.2.1 Target: Creation of Jobs and Growth

The number of data workers is expected to grow in the three scenarios because we expect the use of data-driven innovation to increase even in less favourable economic conditions. The absolute number of data workers is expected to be different by scenario, with total demand of 7.8 million in the Baseline scenario versus 6.4 million in the Challenge scenario and 10.4 million in the High Growth scenario.

Although the indicators of the study do not estimate the new created jobs, we believe that estimating the number of data-related jobs and their trends provides an even better

SMART 2013/0063 European Data Market

measurement of the penetration of data into the production processes. Data is in fact increasingly used by workers with different occupations and functions.

Using and exploiting data is becoming a pervasive phenomenon, which does not exclude other activities. Therefore, what is relevant is the data worker indicator trend — even more so than the new jobs created.

Table 46 EDM Indicators and key policy targets (2)

Key Policy Target	EDM Indicator
Increase the number of data-related jobs (at least 250.000 new data related jobs in Europe in 2017)	Indicator 1 — forecast of data workers growth in the period 2014–2020
Number of European data workers increased in the 2014-2020 period (2014-2020)	Indicator 1 Number and forecast of data workers in Europe Indicator 5 Gap between demand and supply of data workers

Source: European Data Market Monitoring Tool, IDC 2016

Table 47 EDM Indicators and key policy targets (3)

Key Policy Targets	EDM Indicator
Improved use of data for decision-making processes in the private sector and the public sector	Story deliverables (see chapter 8) Data User survey (see par.5.3.4) Citizen's Reliance on the Data Market indicator (see chapter 11)
EU suppliers reaching a market share by 2020 compatible with the size of our economy (30% of world market) (by 2020)	Not available
New advanced methods, systems, technologies, products, services, as measured by the data market monitoring tool	Data User survey (see par.5.3.4)

Source: European Data Market Monitoring Tool, IDC 2015

The improved use of data by businesses and government organisations is not included as a specific indicator, but is extensively covered in the story deliverables, summarised in Chapter 9. Concerning the improved use of data for decision-making processes, we have designed an indicator (indicator 6) able to measure to what extent citizens are taking data-driven decisions in their daily life. Our research found a wide range of possible indicators of the "data maturity" of citizens, but almost none with the required quality and availability.

We have selected an approach based on two proxy indicators, potentially available for all the EU MS and periodically updated, as follows:

- An indicator measuring adoption — the average expenditure in wearable computing per citizen, sourced from IDC
- An indicator measuring the capability of citizens to use data in their daily life — the average level of digital skills measured by the human capital index, sourced from the EC

191

Our assumption is that MS with a high diffusion of wearable computing devices and high level of skills will have a high share of citizens making decisions based on data-driven solutions. We have tested this assumption by analysing a scatterplot combining both indicators, producing valid results. However, we need further work on the adoption indicator to be able to deliver a compound indicator.

Concerning the EU industry market share, the EDM Monitoring Tool analyses all enterprises with a legal presence in Europe.

To measure market shares, it would be necessary to investigate the ownership of these enterprises to differentiate between native Europeans and branches of foreign multinationals. This would have required a different approach to the survey, losing other more useful evidence.

In addition, as explained before in the report, estimating imports and exports in the data market proved to be extremely ambitious: the lack of sufficient data and reliable statistical sources were some of the challenges faced by the measurement.

The current release presents the estimate of the share of the EU data market compared to the worldwide value based on IDC data, for the U.S., Brazil, and Japan.

14 CONCLUSIONS

14.1 The European Data Market: Measuring the Indicators

This report presents in detail the results of the third and final round of measurement of the European Data Market Monitoring Tool, carried out in the period October-November 2016 based on updated statistical sources and market data on the years 2015-2016. The report presents a set of indicators, measuring the European population of data workers, the value of the data market, the number of data user enterprises, the number of data companies and their revenues, and the overall value of the impact of the data economy on EU GDP. All indicators are presented for the years 2013 through 2016 and forecast to 2020, exploring three alternative potential scenarios of evolution: A Baseline scenario, a High Growth scenario, and a Challenge scenario.

Designed along a modular structure, the European Data Market Monitoring Tool was conceived and built around a core set of quantitative indicators to provide a first assessment of the emerging market of data at the present time and with projections to 2020.

The key areas covered by the indicators measured in this report are:

- The data workers and the balance between demand and supply of data skills;
- The data companies and their revenues;
- The data user companies and their spending for data technologies;
- The market of digital products and services ("Data market");
- The data economy and its impacts on the European economy.
- Forecast scenarios of all the indicators, based on alternative market trajectories.

The main findings by key area of investigation are outlined in the following paragraphs.

14.2 Forecasting the Data Market Indicators

The European Data Market forecast scenarios have been updated to take into account new disruptive events, such as the Brexit referendum in June 2016 voting for the UK leaving the EU and the election of Donald Trump to the US presidency, which have greatly increased political and economic uncertainty.

To do so we have leveraged IDC's analysis on the potential impacts of Brexit on IT spending in the EU published in July 2016[50]

---

[50] The Brexit Impact on IT Spend in the U.K. and Western Europe: A Scenario Analysis Insight (Doc #EMEA41570216 / Jul 4, 2016)

14.3 The Brexit Impact on the 2020 Scenarios

In this third and last round of measurements of the European Data Market, we have incorporated the analysis that IDC has recently developed to understand the future developments of the ICT market in Europe and the impact that the on the EU data-driven economy as a whole. In fact, IDC has predicted three possible paths of development following the U.K vote to leave the European Union:

- **Path 1: "Challenging Transition**" — 70% probability. This scenario would see a decline in U.K. GDP at first, but a new relationship set up in some form of bilaterally negotiated agreement in the medium term. Overall we would expect the IT forecast to be revised downwards by more than 2% through to 2020 on a compound annual growth rate (CAGR) basis. Western European IT spend would be expected to remain fairly stable.
- **Path 2: "Disruptive Transition"** — 20% probability. This is most pessimistic scenario and assumes contagion in terms of multiple referenda and immense pressure on the EU model, creating further economic uncertainty. IT spend in this scenario would be expected to decline significantly in the short term and would struggle to rebound in the forecast period in the U.K. and Western Europe. Overall we would expect the forecast to be revised downwards by close to 5 % through to 2020 on a CAGR basis.
- **Path 3: "Swift Transition"** — 10% probability. This assumes strong leadership steps into the existing vacuum and an orderly Brexit process occurs that avoids short-term turmoil and drives economic growth for the U.K. in the medium term. IT spend is affected mildly in the U.K. in 2016, but rebounds quickly in 2017 and beyond. Europe IT spend unaffected.

In the wake of Brexit, based on recent feedback from a number of large enterprise leadership teams, IDC expects a "wait and see" approach as the political and economic lines are redrawn. IT spending will likely shift, but the strategic transition towards the digital enterprise will remain, and in fact is likely to accelerate with a greater focus on cost optimization and IT value to the organization's bottom line. All in all, this materializes in the following three scenarios:

Baseline Scenario

The Baseline scenario is defined by a continuation of the 2015-2016 moderate growth trend of the European economy, creating favourable conditions for investments in digital innovation in general and data technologies in particular. The increasing diffusion of IoT and Cloud Computing will encourage business demand for Big Data technologies, while the nearly universal penetration of mobile and social technologies by 2020 will herald the emergence of a "hyperconnected" society, where consumers will rely on multiple real-time services for their daily life, often supported by data applications. It is also expected that high-speed broadband infrastructures will be available across Europe and will not become a bottleneck for the data market development.

In this scenario, policy will play an important role to support supply, but have a mixed success in promoting demand, an inherently more difficult objective. Policy initiatives will succeed in supporting the growth of the data industry through R&D investments, the support of digital entrepreneurship, and the successful deployment of the contractual Public Private Partnership on Big Data Value (BDVA PPP). The EU will protect trust in the data economy by successfully deploying the General Data Protection Regulation, achieving greater harmonization across the EU and reducing the administrative burden

on businesses. On the other hand, the removal of regulatory barriers preventing the free flow of data cross-borders is unlikely to have effects before 2019-2020. The support of pilot projects and innovation spaces for experimenting with data innovation will help advanced and already interested potential users.

This scenario foresees 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. However, advanced enterprises are a minority of the potential users' population, and in this scenario we foresee only a slow growth of take-up by mainstream, traditional enterprises. For that reason, in this scenario the supply-demand interaction is still strongly dominated by the supply push.

High Growth Scenario

In the High Growth scenario, Europe's economic growth in the next years will be similar to the Baseline scenario, but it will be characterised by a stronger driving role of digital innovation, with higher overall ICT investments as a share of GDP. Solutions combining innovative digital technologies (such as IoT, Cloud and Big Data) will be more widely implemented and more European enterprises will engage in Digital Transformation before 2020. The data market will enter a faster growth trajectory and the adoption of data technologies will spread beyond the minority of pioneers to a wider population of mainstream users. 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. This is a classic virtuous cycle mechanism, which may happen if data technologies take-up starts climbing fast enough to generate momentum. Because of network effects typical of ICTs, rapid diffusion multiplies the benefits for users in their interactions and makes it easier to consolidate standards and interoperability, reducing further the barriers to adoption.

To enable this scenario, we must assume a set of very favourable framework conditions which are able to trigger a faster take-up. First, the adoption of all digital technologies is mutually reinforcing, so we assume a faster pace of diffusion for IoT, Cloud, Mobile as well as data technologies. Second, we must assume a leap ahead of potential benefits' awareness and willingness to adopt data technologies by mainstream users and specifically by SMEs. Third, but not less relevant, we must assume a removal of existing regulatory barriers within the forecast period. In this scenario, policy initiatives will succeed in supporting supply as detailed above, but will also have better success in promoting demand. Policies enabling the free flow of cross-borders data and the re-use of data sets will create positive effects on demand starting from 2017-2018. All the other positive factors described in the Baseline scenario must also be present. As a consequence, the value of the data market and of the data economy by 2020 will be substantially higher than in the Baseline scenario.

Challenge Scenario

In the Challenge scenario, the combination of a less positive macroeconomic context than in the Baseline scenario, less favourable framework conditions, and slower diffusion of digital innovation will combine to push the data market into a low growth development path. This is a fragmented scenario, where the Digital Single Market will fail to materialize before 2020. The supply-demand dynamics will be dominated by the technology push, since the demand pull will be weak. The level of adoption of data technologies by 2020 will be limited to a smaller population of potential users than in the

195

Baseline, as market barriers to entry will remain high. This scenario therefore explores the potential risks and consequences of failing to remove the barriers to the development of the data economy in Europe.

This scenario still foresees an increase of the diffusion of digital technologies such as IoT and Cloud, but at a slower pace than in the Baseline. The dynamics of mobile and social technologies should not be much different in this scenario, given their strong momentum and their closeness to nearly universal diffusion. Therefore, the "hyperconnected" society will become closer in this scenario too, but will be less well developed than in the Baseline or High Growth scenarios. It is possible that the diffusion of high-speed broadband infrastructures across Europe will be incomplete, with the risk of a digital infrastructures divide between and within the Member States. This will be another element of weakness for the development of the data market.

In this scenario, both supply-side policies and demand-side policies will tend to have weaker impacts and to be deployed more slowly in time. Policy initiatives will still succeed in supporting the growth of the data industry through R&D investments, the support of digital entrepreneurship, and the successful deployment of the BDVA PPP, but to a lesser extent than in the Baseline scenario, given the lower propensity to invest by the private sector. Policies addressing enabling conditions, such as the removal of regulatory barriers to the free flow of cross-border data, will be delayed in time and be less effective than in the Baseline scenario. As result, the value of the data market and of the data economy by 2020 will be substantially lower than in the Baseline scenario.

In the wake of Brexit, the three scenarios above have been further adjusted and now incorporates the following changes:

- The GDP forecasts have been reduced compared to the previous scenarios release for the Baseline and High Growth scenarios; the Challenge scenario is the least changed because to some extent it already included negative assumptions. Variations are stronger for the UK than for the other EU27.
- ICT spending is influenced by the macroeconomic uncertainty but also by the combination of opposing trends, the decline of spending in traditional IT and the increase of spending in new technologies which are however more productive and efficient. This leads us to adjust downwards our 2020 forecasts for ICT spending value in 2020 in all scenarios.
- The 2020 data market forecasts have been revised downwards in the baseline and high growth scenario but only by approximately 5% (compared to a minus 20% for ICT spending) since this is an emerging market with strong momentum.
- The 2020 data market value in the challenge scenario was decreased because it already included pessimistic assumptions about overall growth.
- The number of data companies is substantially unchanged in all the 2020 scenarios: this because none of the new developments affect negatively the number of new companies entering the data market as suppliers. In fact, less competition from the UK which included the highest number of data companies might even be positive for some other MS industry.
- The 2020 forecast number of data users has not changed for the challenge and baseline scenarios, but has been reduced in the high growth scenario due to a smaller number of SMEs buying data market products and services. The assumption is that the lower growth of the data market in 2020 will mostly come from lower SMEs take-up (since the demand by large enterprises is expected to be more resilient and driven by the need to compete with their peers).

14.4 Measuring Data Workers

The 2015 data workers measured 6 million in EU28. For the year 2016, we estimate that data workers are increasing with a growth rate 2016/15 which is nearly 2.6%, reaching 6.1 million units. The data workers trend is growing, thanks to the growing trend of the data market.

The average number of data workers per user company (i.e. the number of data workers in relation to the number of data users) is stable, and remains around 9 units per company. A growing trend of data workers with a stable number of workers by user company induce thinking that the diffusion of data products and services is gradually spreading.

As in the previous measurements of the European Data Market monitoring tool, European data workers continue to be distributed in nearly all industries, but their employment share by industry varies substantially. Four industries — manufacturing, wholesale and retail, professional services, and ICT — represented nearly 62% of data workers in 2016 with no significance differences with the previous years (nearly 3,800 on the overall 6,161). In absolute terms, professional services count for 20% of the population of data workers, followed by wholesale and retail with another 18%, and then manufacturing (12%) and information and communication (11%).

14.5 Measuring the Data Worker Skills Gap

This indicator measures the potential gap between demand and supply of data worker skills in Europe. If demand is higher than supply, there is a data skills gap. If supply is higher than demand, there is over-supply and unemployment. The measurement is based on a model estimating separately demand and supply, taking into account the supply of graduates, the level of unemployment, and the entry and exit flows in the data worker market.

According to our model the year 2016 saw an imbalance between demand and supply of 420,000 unfilled data workers' positions, corresponding to 6.2% of total demand – up 0.3% from the same value for the year 2015[51]. Our new estimate for 2016 confirms that the data skills gap in Europe is on the increase, although at a slower pace than in the previous years - the gap almost doubled in 2015 vis-à-vis 2014 and it now grows by approximately 6% in the period 2015-2016.

By 2020, based on the updated demand and supply trends presented in this report, we foresee a continuing imbalance between demand and supply, under all 3 forecast scenarios. In the Baseline scenario the data skills gap will grow at something more than

---

[51] The measurement of this indicator is based on a model built on several assumptions, particularly concerning the share of graduates who choose to become data workers and the entry and exit flows to the data workers labour market. The results should be considered as estimates. The results for the total EU28 are more reliable than the results by Member State. Official statistics from Member States are not available to clarify the specific dynamics of supply and demand by Member State and particularly the potential mobility of workers within Europe. The capture of this information by Member States should be encouraged for future analysis of the skills situation within the European Union.

16% over the next four years totalling a number of unfilled positions of almost 770,000 in the EU28; the High-Growth scenario will see the gap expand considerably at a 60.7% pace with almost 2.8 million positions unfilled; conversely, under the Challenge scenario, the gap will actually diminish with respect to the year 2016 to reach 226,000 units, hence marking a negative CAGR of -14.4%. In fact, according to the Challenge scenario, in 2020 the data worker career will be less attractive for professionals engaged in other careers (restraining one of the main inflows of data workers supply) but also the demand for data workers will be negatively affected, due to a general cooling down of the overall economy. As a result, both the supply as well as the demand of data workers will suffer, with the latter diminishing at a faster pace than the supply. As a consequence, there will still be an excess of demand but at a lower level than in the Baseline or in the High-Growth scenario and the overall gap will be reduced to 226,000 units. This Scenario will present a very varied and divided picture where some of the largest and most advanced EU economies will still experience a considerable gap (this is the case of the UK, Germany and, to a lesser extent France) whereas weaker economies such as Italy and Spain will actually witness an oversupply of data workers.

Fuelled by an intensified demand of data and by higher penetration rates of data-related technology, the High Growth scenario is characterized by a fast increase of the data workers demand which risks to lead to almost 2.8 million unfilled positions, equivalent to a share of 27% of total demand.

In the Baseline scenario, the picture would be similar to what we presented in the previous release of this report although the overall gap will be on the increase, similar to what we estimated for the year 2016. According to this scenario, in fact, in 2020 the overall gap in the EU28 will represent 9.8% of the total demand of data workers – up more than 3% points with respect to our previous estimates.

14.6 Measuring the Data Companies

In 2016 the data industry in Europe featured almost 255,000 companies (up 2.3% from the previous year) representing a share of 14.1% of enterprises populating the ICT and professional services sectors in the same year amounting to slightly more than 1.8 million companies - as a reference, data companies represented a share of 13.8% over a total of 1.8 million ICT and professional services in 2015. The above results confirm that European companies continue their journey towards innovation through the adoption of data-related technologies and the subsequent development of data-based products and services.

As in the previous years, the number of data companies by Member State in the EU in 2016 continued to be heavily concentrated in the two major countries, the U.K. and Germany. The distribution of data companies reflects the concentration by country of the ICT industry. The landscape is therefore dominated by the U.K., followed by the largest EU member states, with a long tail of small groups of data companies in most countries. This distribution is not simply a mirror of the size of each of the member states economies, but is more closely correlated with the presence of a strong ICT industry, and a dynamic professional services industry.

The potential growth trend of the EU data industry in over the years to 2020 is high, driven by the expected high investments in R&D and innovation in data technologies.

- We estimate the number of data companies in 2020 to increase considerably, especially under the Baseline and High Growth scenarios. Under the Challenge scenario, data companies in the total EU could amount to 265,250 in 2020 marking a Compound Annual Growth Rate (CAGR) of 1%. The same CAGR could be at 5% under the Baseline scenario and would reach 8.9% in the High Growth scenario.
- The exit of the UK from the EU will not substantially change the picture but it will exert a somewhat negative influence, especially in the short to medium term. The number of data companies in the EU27 (EU28 less the UK) will grow of 1.9% year-on-year in 2016 as opposed to 2.3% in the same period for the EU28. In 2016 the overall share of data companies on the total of the industry will be lowered of more than 3 percentage points if the UK were not to be included in the calculation (total share of data companies in the in the EU27 would amount to 10.9% in 2016 as opposed to 14.1% in the EU28 in the same year).
- The impact of Brexit would be smoother in 2020 with only very minor variations in the CAGR 2016-2020 under the three scenarios considered – the Baseline scenario would remain unaffected (with a CAGR of 5% over the period in both the EU27 and EU 28; the Challenge and High Growth scenarios would only suffer a minimal 0.1% with a CAGR 2016-2020 of 0.9% under the Challenge scenario in the EU27 vis-à-vis a CAGR of 1% in the EU28 and a CAGR of 8.8% in the EU27 vis-à-vis a CAGR of 8.9% in the EU28).

14.7 Measuring the Data Users

Data users are organizations strongly relying on data to their objectives. In other words, they generate and exploit their own data, collect online customer data intensively, submit this data to sophisticated analysis and use them (and the results obtained by them) to improve their business. As such, data users represent the demand side of the data market.

In 2016, data users in the EU28 will reach more than 661,000 units according to our latest estimate, thus growing 1.6% year-on-year and representing an overall penetration of 6.4% over the 10.3 million potential user companies – a persisting low penetration across the four years under consideration (6.2% in 2013 and 6.3% in 2014 and 2015).

As opposed to data companies, data users can be found in every industry sector and are therefore less concentrated than the former, also in terms of Member State distribution. Still, the UK leads the way in this respect too, followed by Germany, Italy and Spain, which muster alone almost two thirds of the total number of the European data user companies. Some countries continue to exhibit growth rates in 2016 that are way above the EU average: this is the case of the Netherlands, the UK, Sweden, Cyprus, the Czech Republic and, to a lesser extent, France and Germany, thus confirming the positive trend already displayed in 2015.

In line with the results of our Second Interim Report, the number of data user companies is expected to grow steadily over the period 2016-2020 with a CAGR ranging from 0.3% under the Challenge scenario to 13.5% according to the most favourable scenario.

At Member State level, the concentration of data users will be lower than the concentration of data companies. The UK and Germany will continue to have the lion share of data users under all three scenarios: together with the rest of the largest EU

199

economies (France, Italy and Spain), they will total more than 70% of all data users in 2020 in the Challenge, Baseline and High Growth scenarios. This trend will be particularly pronounced under the latter scenario, which is characterized by a marked take-up of data-related technologies boosting, in turn, the overall population of data users in Europe.

14.8 Measuring the Data Market

The marketplace where digital data is exchanged as "products" or "services" as a result of the elaboration of raw data is what we have defined as data market in this study. The European data market in the EU28 is now estimated at EUR 54,351 million in 2015 and at EUR 59,539 million in 2016, thus exhibiting a solid year-on-year growth of 9.5%. This positive trend will continue throughout the next four years, although at different paces according to the selected scenarios, registering a 7.5% growth rate under the Baseline scenario, a 15.7% under the High Growth scenario and a 4.3% under the Challenge scenario. If we exclude the UK from our projected estimates, the value of the data market will grow at the same pace in the EU27, although slightly slower in both the Challenge scenario (with a CAGR of 4.1% vs. a CAGR of 4.3% in the EU28) and in the High-Growth scenario (with a CAGR of 14.9% vs. 15.7% in the EU28). As one of the most vibrant ICT-driven economies, the UK confirms its importance in an economy which is more and more driven by data and data-related products and services. As a result, its exclusion will have a negative (although minor) impact on the overall dynamics of the European data market.

As a percentage of the total ICT spending in the EU28, the data market is now estimated to represent a share of 9.5% - against a share of 8.8% in 2015 – and is expected to improve under all of the three scenarios considered in this study with respect to our previous estimates (12.5% under the Baseline scenario; 13.1% under the Challenge scenario and 14.5% in the High Growth scenario). The reason for this positive development is explained by an overall ICT spending as a whole that is likely to grow at a more moderate pace in the years to come, while the data-related components of its spending (i.e. the data-driven technologies such as Big Data and analytics) will increase at a faster pace.

By industry, Europe's data market continues to be dominated by sectors that make a significant usage of data-related technologies, i.e.: manufacturing, finance, the area of professional services and the ICT sector. As in our previous analysis, the most dynamic growth by industry in 2016 is performed by vertical markets that still hold a relatively small size of the overall data market spending by industry. This proves that the data-related technologies are rapidly finding new ground in previously unchartered areas and are growing fast in sectors like home, healthcare activities, education and the public sector as whole.

14.9 Measuring the Data Economy

The economic impact analysis is an effective tool for the scrutiny of the impacts of a multipurpose and widespread innovation such as the data products and services. This kind of approach helps subdivide the impacts in order to better understand the source of such impacts, and whether they can be considered new additional impacts on the economic system. An in-depth and reliable analysis focusing on impacts should be based

on specific field research, but the survey provided interesting insights about the impacts gathered by the companies adopting data products and services. A significant share of the companies is starting to see benefits in terms of additional revenues.

The overall value of the data economy grew from the EUR 247 Billion in 2013 to the EUR 285 Billion in 2015, almost reaching EUR 300 Billion in 2016, marking a growth rate year-on-year of 5.03% in the period 2015-2016. According to our estimate the value of the data economy in 2016 was worth nearly 2% of the European GDP. In 2013 the total impacts estimated for data products and services represented EUR 246,840 Million, which is equivalent to 1.83% of the EU GDP; in 2014 EUR 257,589 Million, which is equivalent to 1.85% of EU GDP. In 2015, we estimate the same indicator to be at EUR 285,633 Million, representing 1.94% of the overall EU GDP.

The impacts gathered by the user industries (forward indirect impacts) represented almost 56% of the total impacts in 2013, approximately 54% in 2014 and grow slightly to 55% in 2016, corresponding to 1% of total EU GDP. When compared with the total impacts in 2013, we can see that the total impact is not increasing very fast and significantly year on year. This is quite normal and in line with the overall impacts of innovations. First of all, because the penetration of the data products and services is not so fast and also because the growth rate for the accumulated impacts may not be as fast as the growth rate for the data market. The scenarios at 2020 show that a high penetration in the user industry produces relevant and fast impacts in terms of GDP.

The High Growth scenario, under similar macroeconomic conditions, produces relevant impacts on the user industries. A fast penetration of data products and services produces relevant effects in terms of GDP though the benefits achieved by the user industry. This means that policies that leverage on increasing demand for products and services may provide relevant impacts.

14.10 A Complementary View of the Data Market: Real-Life Stories

The data market is already having an effect on a number of industries and sectors. Our analysis has focused initially on five key segments, covering the public sector as well as the retail, manufacturing, utilities, and banking industries. For this Second Interim Report, we have also included a series of thought-provoking considerations from the precision agriculture sector, as well as from a wider and far-reaching debate around data-ownership and the use and re-use of data in several economic sectors.

An initial attempt to categorise data-related benefits suggests we rearrange these benefits along five main categories: revenues, costs, operational efficiency, organisational effectiveness, promotion of ICT adoption, entrepreneurship and new ventures.

- Data-driven technologies are helping European organisations to increase revenues. This is particularly evident in customer-facing activities where the use of Big Data and analytics is proving to be fundamental in finding new customers or reactivating existing ones (as in the retail industry), or in better understanding their needs and requirements (utilities, banking, retail industries) and generating new turnover (banking industry). As an example, Morrison's, a British retailer, increased its number of new or reactivated customers of 150% in 2014 and, all in

all, 20% to 25% increase of overall return for an electricity supplier in Europe through better customer relationship, lower cost to serve via digital channels, direct marketing and improved behavioural demand-response.

- Big Data and analytics can also drive significant cost optimization. The retail industry is capitalizing on new, more efficient stock management practices offered by a more sophisticated interpretation of existing data, while manufacturers benefit from lower levels of scrap and rework. We found, for example, that Tesco, a British retailer, achieved yearly savings of £6 million through efficiency in stock management (i.e. reducing discounts in cases of over stocking) thanks to the adoption of Big Data and Analytics technologies.

- Operational efficiency is attained in banking through a considerable reduction in the time spent on trouble-shooting, administrative practices, and risk-containment activities (such as fraud detection). Government agencies can look at severe budget cuts more comfortably as data-driven technologies enable them to execute complicated data cleaning and data validating operations more swiftly and with fewer resources. To quote another example, Nottingham City Council reduced time for handling information from 3 months to 3 hours by matching information sources automatically and not by hand.

- Big Data and analytics can increase effectiveness in utilities and manufacturing and improve key performance indicators such as order-delivery time, response time, and time to market, while the public sector can devote its (often scarce) resources more effectively toward prioritised policy needs thanks to better use of data analytics. For example, our research demonstrated that the manufacturing sector could obtain an overall reduction in order delivery time from 3.5 to 0.8 days for the average European manufacturer by improving sales and operational planning through the adoption of last generation Big Data and Analytics technologies.

- Data-driven technologies also allow traditional, non-ICT driven sectors to significantly upgrade their uptake and deployment of advanced technologies with positive impacts on all the above mentioned categories of benefits. As an example, Big Data and Analytics can be applied to the agricultural sector giving birth to what is now called "Smart Agriculture" or "Smart Farming". Our research showed that SMEs active in the agricultural sector and other small farmers with a traditional low usage of ICT, can reap up to 25% savings in costs due to a reduced use of fertilizers, increase productivity of at least 5% thanks to crop increase, or, again, augment crop yields through smart water management and smart fertilization practices.

Additional benefits from enhanced data sharing and data exchange

In parallel with the role of data-related technologies in supporting the data market and exerting positive impacts to the European data economy in general, the study team also addressed additional elements directing affecting the very availability and usage of data – the concept of data ownership and its related issues; the role of data marketplaces and the presence of industry data platforms as well as the role of data skills to accompany the current digital transformation.

Data Ownership and its related Issues

The way data ownership and data access are managed and regulated can directly affect the functioning of the data market. Companies having a high concentration, or accessing

huge amounts of data, could easily incur in situations of market asymmetry, which – in turn – may result in different forms of market distortion. In our case studies we found no significant evidence of severe market abuse as the current level of data exchange and data re-use does not seem to cause stark hindrances to the overall market efficiency, at least at this stage of the process. As a result, data ownership could be considered within the broader framework of growth, innovation and competition policies and not seen simply as a contractual issue or a legal matter. This is not to say that a certain number of guidelines, as well as new types of model contracts, could be fruitfully developed by the industry to help data-stakeholders come to terms with emerging business models and new business cases.

The Role of Data Marketplaces

The recent developments of the so called "3rd Platform of computing"[52] – internet-centric computing systems exploiting the power and capabilities of four key technology developments (cloud, mobile, social and big data analytics) have significantly accelerated the exchange of electronic data giving birth to the phenomenon of (electronic) data marketplaces. In their simplest form, data marketplaces are online stores where people can buy and sell data, but they are evolving towards a more sophisticated intermediary role, potentially central to the emerging data ecosystem. They can maximize the value of data by facilitating the exploitation and re-use of proprietary data as well as its integration with open data, by providing a platform for data holders, service providers and data users to work together, and by building trust in the data value chain, thanks to their role as independent third party. By doing this, data marketplaces play the role of multi-sided digital market platforms, where the value of the products exchanged is multiplied by the interaction of the platform users. Data marketplace platforms of this kind can respond to a variety of emerging data market needs, by improving the efficiency of the data value chain, reducing transaction costs, providing a platform for sharing and re-using data sets, and solving data interoperability, privacy and security problems on behalf of data holders who may not have the necessary skills.

The presence of Industrial Data Platforms

Data marketplaces are not the only way organizations regularly exchange and share an increasing amount of data. Other forms of initiatives are emerging across the two shores of the Atlantic. Industrial data platforms, in particular, are rapidly becoming a trusted network for data access, transfer and usage, thus disrupting traditional value- and supply-chains and bringing together a wide array of disparate players from a multitude of sectors. However, at this initial stage, industry data platforms are not easy to define with their actual implementation displaying a diverse picture where two opposite models coexist. On the one hand, industrial data platforms may take the form of open, multi-company-led environments that are conceived to meet the requirements of a wide community of industry users from different industry sectors; on the other hand, single-company initiatives are emerging where an individual organization (usually a prominent ICT player or a leading industry actor) establishes its own data platform and open it to other companies for commercial purposes and primarily within the boundaries of a specific industry sector.

---

[52] http://www.idc.com/prodserv/3rd-platform/

Companies are already starting to realize the advantages of this new model of exchanging data in terms of both productivity and efficiency improvements. IDC Manufacturing Insights, for example, foresees that by 2019, 75% of manufacturing value chains in EMEA will undergo an operating model transformation with digitally connected processes that will improve responsiveness and productivity by 15% on average. Again both the Boston Consulting Group and PwC believe that digitization and interconnection of products and services will also allow companies to again additional revenues of 2% to 3% per year on average. If confirmed, these gains could amount to approximately €30 billion per year for Germany's industry alone and to €130 billion per year for the European industry as a whole

According to IDC, 65% of large enterprises worldwide will have committed to becoming information-based companies in 2016, shifting the focus from resources, labor, and fixed capital to relationships, people, and intangible capital[53]. Understandably, new digital skills are required to reap the benefits of such shift but most of the EU companies and organizations do experience a skills mismatch issue: European employers are struggling recruiting people with the skills needed, unemployment remains at high levels and over 25% of young adult employees are overqualified for their jobs (Eurofound, 2014 and Cedefop 2015). This bleak picture is confirmed by the results of the European Data Market tool[54] : its Data Skills Gap indicator projects a potential supply-demand gap of approximately half a million data workers' jobs in the EU by 2020 under the baseline scenario. As a result, only 64% of European firms expect their digital transformation objectives for 2016 to be fully reached (European Commission, 2016; The Economist Intelligent Unit, 2016) and many companies are bound to consider new and more compelling initiatives to upskill, re-skill, or simply acquire newly hired employees with the necessary skills.

Data Skills and the Digital Transformation

In terms of managing the needs for new digital skills, our analysis across European organizations highlights the importance of bringing about an all-encompassing process of cultural transformation in parallel with the digital transformation journey. In other words, the demand for highly sophisticated and specific data skills remains limited and not too difficult to meet through the acquisition of a small number of highly specialized data scientists. On the contrary, upskilling the digital skills of the overall existing workforce poses a much bigger challenge and requires a deeper change. When it comes to the training and upskilling of the workforce, our primary research shows a clear preference for internal training: it is less costly, it allows the process of digital innovation to be spread and steered by the company according to its chosen strategy and reduces the risks related to the exposure of business-sensitive information to external third-parties. As for the sourcing of data skills, European companies seem to favour the route of external acquisition through the hiring of young graduates with very specific data skills.

---

[53] IDC Digital Transformation FutureScape 2016: Worldwide Predictions, IDC 2016

[54] See: The European Data Market study (SMART 2013/0063): Second Interim Report, June 2016. Also in: http://new.datalandscape.eu/

From the above, a few critical issues requiring policy-makers' attention emerge:

- The need to deal with employees with obsolete skills who are not willing or able to adapt to digital skills. Our research points to a widespread lack of awareness by enterprises and by employees about the depth of the cultural transformation currently underway.
- The need to improve the visibility of data skills recruitment challenges to help companies and organizations find the right people with the appropriate sophisticated data analytics skills;
- The need to accelerate the current industry-university collaboration to increase the number of data-related internships, their duration and their scope so to ensure that data competences are available and up-to-date.

14.11   Measuring the Citizens' Reliance on the Data Market

This indicator was conceived to measure the level of citizens' reliance on data and to provide a more complete picture of the importance and social benefits of the data economy to the EU. It aims to complement the "business orientation" of the other indicators, with the aim to provide a snapshot of how citizens are taking advantage of data-driven solutions in their daily life. It is still a mainly "experimental" indicator, opening new ground and suffering from the lack of relevant and systematic data.

Our research found a wide range of possible indicators of the "data maturity" of citizens, but almost none with the required quality and availability. We have selected an approach based on two proxy indicators, potentially available for all Member States and periodically updated, as follows:

- An indicator measuring the adoption, that is the % of citizens owning a wearable computing device, sourced from IDC;
- An indicator measuring the capability of citizens to use data in their daily life — the average level of digital skills measured by the human capital index, sourced from the EC.

Their combination is our Citizens' data indicator which is measured as follows:

- % of citizens owning a wearable device (calculated by dividing the number wearable devices sold in last 2 years by the n. citizens) * % of citizens with basic ICT skills.

The indicator is based on 17 MS out of 28 (excluding the Baltics, Bulgaria, Romania, Luxembourg, Croatia, Slovenia, Slovakia, Malta and Cyprus). These MS correspond to 92% of the EU population.

The results show that in 2016 only 4,1 % of EU18 population used data provided by wearables to drive their decisions, varying from 10% in U.K. to 0.2% in Romania.

While these data are relatively low, this is a very new phenomenon and the expected growth is strong. Based on IDC market prediction, we estimate that by 2019 more than 20% of EU citizens use data to take decision on a wide array of daily activities such as doing sport, dieting, commuting, holiday-making and entertainment, just to mention a few. The U.K. is at the forefront of adoption of this technology. Eastern European and Mediterranean Member States are at the bottom of the ranking. Even more worrying is that the countries with higher adoption in 2015 are those who grew faster, hence widening the gap.

As per 2016, these data are likely to marginally underestimate the phenomenon as the skills of the early adopters are expected to be more advanced and the indicators do not capture the diffusion of health apps, for which reliable data are not available.

Worldwide Monitoring of the Data Market

According to our third and last round of monitoring, the European data market and economy in the period 2013–2016 was consistently second to the U.S. in value and, to a lesser extent, in growth. In terms of impact of the data market on GDP, however, in the same years both the U.S. and Japan had higher levels of incidence. In each of these countries the dynamics of the data market are positive and the growth outlooks are promising.

- The total number of data workers in the U.S. and their share of total employment are almost twice as high as in the EU, with penetration rates close or above 7% throughout the period, versus EU penetration rates of approximately 3% during the same years. This is clearly a result of greater data-driven innovation in U.S. enterprises and a more mature data market.
- The growth rate for the data worker population is four times higher in U.S. than in Europe in the period in 2016. The data worker population in Japan is also quite high at over 3.7 million in 2016 and the share of total employment is also higher than in Europe at 5.8%. This reflects the high level of IT adoption in Japanese businesses.
- Brazil has shown a moderate increase of the total number of data workers in 2016, being now 1.1, with penetration of total employment at around 2% and the growth rate in the last two years was considerably lower than in Europe. *International Comparison of Data Market Value*

A better relative assessment of the level of development of the data market is shown by the comparison of the share of the data market value on total ICT spending. The data market is one of the most dynamic components of ICT investments and a growing share is a good proxy of the fast adoption of innovation.

- In the U.S. the relative weight of the data market on the overall ICT spending increased steadily from 2013 to 2015, passing from 9.8% to 12.3%; however, the estimate for 2016 indicates a minor decrease to 10.9%.
- In contrast, Japan exhibited a more varied picture reflecting the country's economic trend – the data market represented 9.2% of the overall ICT investments in 2014 (down 0.3% vis-à-vis the previous year) but it rebounded in 2015 achieving a noteworthy 11% according to our latest estimates; in 2016 our estimates show a very moderate contraction to 10.9%.
- Notwithstanding its difficult economic and political situation, Brazil managed to maintain a growing trend of its data market size in comparison to the overall country's ICT spending, which in 2016 reached 6.0% and almost doubled the percentage presented in 2015 (3.8%). With a data market amounting to a 9.5% share of the overall ICT investments in 2016, Europe still lags behind both the U.S. and Japan.

We have used the direct impacts (of the data market) and the backward indirect impacts (of the data industry) as a percentage of GDP in the period 2015-2016 to carry out a comparison of the data economy between the EU28 and the three countries taken into consideration in this chapter.

The results are quite interesting: in 2016 Japan took the leadership with a significant 0.9% of GDP generated by the data market in 2016, followed by 0.78% for the US.

SMART 2013/0063 European Data Market

Europe still presents an incidence of economy of approximately half the one of the U.S. On the other hand, Brazil, which quite well over the past three years, shows a contraction in 2016. Brazil decreased the incidence of its data economy of 0.16 percentage points in 2016.

All in all, the above landscape reflects the different levels of maturity of these economies, as well as their stage of data-driven innovation. What is more, the results of both this and the previous round of measurement of the international indicators, confirms the strong potential of the data economy, together with its ability produce tangible benefits for Europe if a rapid development of the data ecosystem is adequately supported.

METHODOLOGY ANNEX

Forecasting Scenarios Methodology

The following tables show the assessment of the data-driven economy strategy actions one by one, in terms of their potential level of impact on the market development (based on our assessment) and the level of uncertainty of outcomes in the period 2014-2020. The objective was to identify the policy actions with high impact and uncertainty, leading to differentiating assumptions for our scenarios. The same exercise was carried out for the DSM strategy actions relevant for the data market.

Table 48 Main Data-driven Economy Strategy Policy Actions by potential impact and uncertainty

Community building	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
1. A European Public-Private Partnership on Data	High	Low: the BDVA cPPP has been launched and has recruited a high number of participants
2. Digital entrepreneurship and open data incubator	Medium – the most relevant market impacts are likely to materialise after 2020	Low: high likeliness of implementation. No differentiator between the scenarios
3. Developing a skills base	High	High. The initiatives are likely to be implemented but whether they will succeed in training a relevant number of skilled data professionals making a difference for the market is uncertain.
5. Identification of sectorial priorities for R&I	Medium – the most relevant market impacts are likely to materialise after 2020	Low: the pilot projects will be launched in 2016. No differentiator between the scenarios
Developing framework conditions	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
Availability of data and interoperability		
1. Fostering Open Data policies	High	High: uncertain success of the policies in terms of adoption and take-up of guidelines and services
2. Data handling tools and methods	Medium – the most relevant market impacts are likely to materialise after 2020	Low: high likeliness of implementation. No differentiator between the scenarios
3. Supporting new open standards	High (see also DSM roadmap)	High uncertainty about potential success (difficult process requiring private initiative)
Enabling infrastructure for a data-driven economy	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
1.Cloud computing	Medium	Low uncertainty: process already started by

SMART 2013/0063 European Data Market

		the ECP. No differentiator between scenarios
2.E-infrastructures and High Performance Computing	High	Low uncertainty, high likeliness of implementation coupled with high demand
3.Networks/ Broadband /5G/	High, but impacts will materialise after 2020	No differentiator between the scenarios – long term action
4.IoT	Medium – the most relevant market impacts are likely to materialise after 2020	Low uncertainty, high likeliness of implementation. No differentiator between the scenarios
5.Public Data Infrastructures	Medium – public data infrastructures are important but insufficient for market development	Medium: complex process, difficulty in coordinating effectively all stakeholders and insuring easy access and usability
Regulatory issues	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
1. Personal data protection and consumer protection	High	High: strong differentiator between the scenarios depending on the implementation process
2. Data-mining	Medium	Medium: initial exploration, actions still to be defined
3. Security	High	High: complex process, just started, uncertainty about timing and take-up by private sector
4. Ownership/transfer of data	Potentially high, but emerging issue driven by new business models	High uncertainty about timing and impacts (see also DSM roadmap)

Source: IDC elaboration on EC Communication "Towards a thriving data-driven economy"

Table 49 Main DSM Strategy Policy Actions by potential impact and uncertainty

Annex: Roadmap for completing the Digital Single Market Actions	Aspects Relevant for EDM	Level of impact on the data market development	Level of uncertainty of potential outcomes, 2020
Key Action 1) Better access for consumers and businesses to digital goods and services across Europe	Cross-border online purchases of digital content, cross-border data mining	Medium	Medium – complex process
Key Action 2) Creating the right conditions for digital networks and services to flourish	Review the e-Privacy Directive (2016)	High, but main impacts after 2020 – no differentiator	High – process not started yet
Key Action 3) Maximising the growth potential of the Digital Economy	Initiatives on data ownership, free flow of data (e.g. between cloud providers) and on a European Cloud	High	High uncertainty about timing and impacts
	Adoption of a Priority ICT Standards Plan and extending the European Interoperability	High	High uncertainty about actual impact

	Framework for public services	on market

Source: IDC elaboration on EC Communication "A Digital Single Market Strategy"

Measuring the Data Workers

Data workers: definition and discussion

Data workers collect, storage, manage and analyze data, as their primary activity: they should be able to work with massive database and with emerging database technology. Data workers are included in the category of the knowledge workers and specifically in the category of the "codified" knowledge (Lundavall and Johnson, 1994); data workers specifically deal with data while knowledge workers deal with information and knowledge. Data entry clerks' primary activity is related to data, so they could be considered data workers; however, data entry is a very routine task and for the sake of this study, data entry clerks are not going to be considered as knowledge workers. Another specific 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.

Within the broader category of data workers, we include the category of **Big Data analysts.**

Big Data analysts 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. Big Data analysts 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).

Big Data analysts are part and parcel of the data workers for the following reasons:

- Because a valuable system of indicators has to consider not only the definitions and indicators which are useful and feasible today, but also the indictors feasible when the industry will overcome the very initial stage
- Because one of the next indicators relate to the skills gap. Again, the skills gap will be calculated with reference to the data workers in general. Nevertheless, when we will discuss the skills gap it will be important to be aware about the skills needed and this can be explored only keeping the skills specifically referred to the Big Data analysts, not only to the larger category of data workers.

Big Data analysts are not going to be estimated but it is important to be aware of these necessary skills.

Technology is an enabling factor which is transforming generation and use of data as well as the related added value (OECD, 2013). Economic and social activities in fact have long made use of data. In recent years, the development path of ICTs increasingly enabled the economic exploitation of data. Technology proves to be important along the data value chain to:

210

- Decline costs,
- Increase the generation and use of data
- Accelerate migration of socioeconomic activities to the Internet with a wide adoption of e-services.

Based on this definition, data workers refer to both supply and demand of data products and services. In one case they deliver data products and services and in the other case they are users of those data products and services for example to take decisions into their enterprises.

Data workers and Big Data analysts are therefore workers using the IT technology to create value from raw data which are available on the market. Data workers create value because:

- They are employed on the supply side and they create value through the sale of data products and services;
- Data workers are employed on the demand side and they create value because they use of data products and services to improve competitiveness and productivity of their companies.

The use of data is pervasive and has penetrated every industry and business function, and data are now relevant production factors, with labour and capital.

Statistical definition of the data workers

Data workers are not classified as such into any of the labour and occupation statistics. As it usually happens, the emerging sector and industries and the related variables are usually not traceable into consolidated statistics. This means that, to estimate such variables, we need to trace the indicators we are interested in, into more general data and to define and find out an approach to estimate them.

In order to define statistically the data workers, we have adopted the International Standard Classification of Occupations (ISCO-08). Clearly, into the ISCO classification, we don't find any category referring to data workers and Big Data analysts. Nevertheless, we can define in which categories of the ISCO-08, data workers and Big Data analysts may be classified. In this paragraph we present the categories of the ISCO-08 where data workers may be classified and counted.

In Annex 2, the detailed table with the list of the ISCO-08 codes selected is presented.

The criteria adopted for the selection of the ISCO-08 codes are the following:

- We have selected the occupations where data workers 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 workers 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 conceptualize 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 workers; the occupations where the data workers are a very marginal part of the

211

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

- We excluded all the data workers 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 workers but since their activity is a routine one and as such excluded from the knowledge economy, they are not considered data workers).

The selected codes finally include:

- 4 major groups (1 digit)
- 9 sub-major groups (2 digit codes)
- 21 minor groups (3 digit codes)
- 52 unit groups (4 digit codes)

The relevance of codes including data workers is shown in the below table: 4 out of 10 major groups (1 digit) include data workers while at the very lowest disaggregation level, 12% of the units include data workers. This represents the perimeter to be assessed.

Table 50 ISCO-08 structure and data workers

	ISCO-08 structured classification			
	Major groups (1 digit)	Sub-groups (2 digits)	Minor groups (3 digits)	Units (4 digits)
Number of codes ISCO-08 structure	10	43	130	436
Number of selected codes including data workers	4	9	21	52
Share of data workers codes in the ISCO-08 structure	40%	21%	16%	12%

Source: IDC elaboration on ISCO codes

In this Final Report we confirm the definition of data workers as adopted in 2014 and that we further applied in 2015. The definition of data workers adopted in this study reads as follows:

Data workers collect, store, manage, and/or analyse, interpret, and visualise data as their primary or as a relevant part of their activity: they should be able to work with massive databases and with emerging database technology, and with structured and unstructured data, to elaborate and visualise data in order to use them for the interpretation of facts and as factors for decision-making processes.

The identification, definition and measurement of data workers is a recent field of investigation. High-tech systems created a massive quantity of data; to use and exploit such data, data analytics professionals are necessary. The data industry is an emerging industry and the use of data is at an early stage. The amount of data which is being produced and used is progressively increasing and the technology and the tools to

process this data are becoming more and more available. The diffusion and adoption of data depend on two factors:

- The availability of technology and tools to collect, process, and analyse data, and make data usable.
- The awareness and capability of users to process and analyse data. It may be that some users, aware of the potential effects of data, analyse it and use it although they "under-use" the technology for data processing and analysis.

In our definition, data workers 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 workers belong to the category of knowledge workers and specifically "codified" knowledge workers (Lundavall and Johnson, 1994); data workers specifically deal with data while knowledge workers deal with information and knowledge.

This is clearly a broad definition of the workers dealing with data. This was a very considered approach. The reason why we decided to adopt such a definition depends on two aspects:

- a- As said above this is a very recent field of investigation and the definitions of the workers dealing with data is in the making
- b- Since the data industry is pervasive and its products and services are going to penetrate most of the industries, we wanted to catch most of the data workers, in whatever industries and with whatever occupations.

Discussion and comparison with other definitions

The recent OECD report Data-Driven Innovation (2015) underlines that data skills are key enablers of data-driven innovations. In fact, a survey conducted by OECD showed that the enterprises using skills related to data and analytics are more likely to innovate and to gain faster productivity growth. This suggests that there will progressively be more and more demand for data and for data skills from businesses from all industries.

The point is that, currently, there is not an agreed definition about data workers and data skills. Specifically, the OECD Report considers the big data area while we do consider a broader area of data, as explained in the Methodology Report.

OECD provides a definition of the "data specialist professions", which is conceptually in line with our definition of the data workers (OECD, 2015, chapter 6). Nevertheless, our approach is a broad approach where we tried to include and to estimate all the data professionals (data workers), while the OECD approach is an extremely narrow approach.

OECD defines data specialists as

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

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

This different approach in the definition is probably related to the fact that OECD aimed at counting the data specialists, based on official statistics" while we aimed at estimating

what we called the broader category of data workers. Official data are not available for the broader category as defined in this study so that it needs to be estimated.

OECD included only the occupations where all the professionals can be defined data specialists and therefore counted with official statistics. IDC, instead, included all the ISCO occupations potentially including data workers; within each of these ISCO categories we estimated (based on a methodology explained in the Methodology Report and in the first Interim Report) how many of the workers counted in the selected occupations may be defined as data workers.

Basically, this means that OECD counts only the ISCO occupations where we are sure that all workers counted are working with data and this is definitely the main part or their job. In this way, OECD excluded many occupations where there may potentially be a share of workers for which data is a relevant part of their job.

The IDC estimate aims catching nearly all the data workers in the European industry.

It is obvious that the data specialists as defined by OECD are a much lower share of employment where compared with the IDC share of data workers. In 2013, OECD estimates that the ratio of data specialists on total employment goes from 0,1% in countries such as Portugal, France and Turkey to over 0.6% in countries such as the Netherlands, the United States, Australia while this ratio reaches 1.6% in Luxemburg. IDC estimates that the share of data workers in EU28 was 3.0% in 2013; in France and Portugal this share was around 3% while in Luxemburg it was 5.2%.

These results look quite consistent. IDC approach estimated the number of data workers within 21 ISCO minor codes (3 digit codes) while OECD counted the workers of 2 ISCO minor codes.

The OECD report also analyses the so called data scientists. If there is not a shared definition for workers dealing with data, the definition for data scientists is even more confused.

Data scientist is a term used by several authors, but unfortunately everyone uses this term with a different definition.

"What data scientists do is make discoveries while swimming in data" (Davenport, 2012); to do so they combine a lot of different skills of software programmers, statisticians, domain experts, and storytellers. Data scientists often need to integrate data from a variety of sources and to build new algorithms which are based on big data specific software systems. Some authors stress that data scientists need to have a very specific domain competence, commercial awareness and knowledge of business processes (Brynjolfsson and McAfee, 2012). Some authors focus a lot on the domain knowledge of the data scientists; nevertheless, we believe this is not something specific for the data scientists. This is something requested and more and more relevant for example for the ICT professionals while the specificity of the data scientists are the technical skills.

The data scientist definition remains somewhat controversial; OECD stresses it is not clear whether this is a new job category and what are its specific skills; in any case, OECD states it should be qualified as the almost talented category of workers dealing with data.

214

In our First Interim Report we defined the data scientists as the data workers dealing specifically with big data. For the sake of clarity and in order to avoid interpretations, we are going to adopt, from now on, a different terminology.

We are going to keep our definition of data workers as the one discussed last year. Instead, we are not going to talk about data scientists anymore but of Big Data analysts.

In this study we have given a broad definition of data workers. What is certain is that there is a category of workers expected to have sophisticated technical skills which have been in rare supply until now.

The ongoing European project named European Data Science Academy (http://edsa-project.eu/) provides a complex definition of the data science, being a person "unique mix of strategic and practical skills, spanning maths and statistical knowledge, machine-learning, domain expertise, data and advanced computing, visualisation and scientific method". Nevertheless, the European Science Academy also underlines that such a mix of skills is very difficult to find in just one person. This is what was also discussed and explained in our first Interim Report. As a consequence, we suggest to talk about data science but not about data scientists, since all the requested skills cannot design only one job profile. This is why a more general approach based on data workers may be (at least for the moment) more appropriate especially when the aim is to estimate the number of workers dealing with data, with more or less sophisticated skills.

In conclusion, our definition of "data workers" is more comprehensive than the OECD one, which is limited to 2 professional categories and does not capture the full range of skills needed for data-driven innovation. The EDSA project definition of data scientist is closer to our "data worker" definition but more oriented to scientific and highly sophisticated skills, less focused on business needs. For this reason and to avoid confusion we will not use the term data scientist in this report, accepting EDSA's definition for it.

Data Worker Share on Total Employment

For the calculation of the data worker share on total employment we used the data reported in the table below which is sourced from Eurostat Labour Force Surveys, aggregating the following occupational categories for the 11 industries selected by the study.

Table 51 Occupational categories selected by the study

OC1 — OC2 — OC3 — OC4 — OC5 — OC6 — OC7 — OC8 — OC9 — OC0

Total employment by Industry	2013	% on total
Mining, Manufacturing	34,016	18%
Finance	6,399	3%
Professional Services	21,732	11%
Wholesale, Retail	40,205	21%
Information & Communication	6,188	3%
Public Administration	14,941	8%

SMART 2013/0063 European Data Market

Total employment by Industry	2013	% on total
Transport and Storage	11,115	6%
Utilities	3,099	2%
Healthcare	23,023	12%
Education	16,063	8%
Construction	14,956	8%
Total	191,736	100%

Source: IDC elaboration on Eurostat Labour Force Survey

Data Worker Estimate Process

First step: the first step is the calculation of the so-called perimeter comprising the industries and occupations where employees may be present with varying levels of intensity: these are the data worker candidates. Based on IDC's elaboration of Eurostat data, the perimeter of workers who are data worker candidates is equivalent to 52 million workers: in 2013 there were in Europe about 52 million workers among which we could find data workers. To calculate the perimeter, we crossed the employment data by occupation (ISCO-08) and the employment data by countries and industries (Eurostat, Ilostat). To do this we had to complete the time series provided by the statistical sources. Where country data for 2013 was missing, we estimated it to get a complete data set for the estimates of data workers in 2013.

Second step: the second step is the estimate of percentage of data workers within the perimeter of data worker candidates. To calculate the coefficients for the calculation of such %, we set the assumptions set above. Then for each industry, based on the available literature and on the survey results, we estimated the proportion of data workers by industry and country.

Third step: the third step is the estimate of data workers active in 2014. Statistical data for 2014 is not yet available, so the estimate for the last year is a forecast. This was based on the following assumptions:

- Countries investing in data products experience an increase in TFP growth
- Data market growth is a function of data labor growth, stock capital growth, and TFP growth using a production function approach
- The data market total factor productivity is similar to the average total European factor productivity
- The ICT capital stock is equal to the sum of the last three years in ICT investment

Methodology of the Data Workers Forecast

The approach for the estimates of data workers in 2020 is the same used to estimate the number of data workers in 2014 and then renewed in 2015 and 2016. Countries and industries investing significantly in data products and services will experience an increase in the overall efficiency of labor and capital or in total factor productivity (TFP) growth. Industries receive productivity gains from data use over the labor productivity gains

received from investments in ICT, for instance because of an improvement in production processes.

We have presented in the §4.2 the methodology for the estimate of the data workers in 2014 which is the same approach we used for the forecast in 2020.

The only difference introduced in the second and third rounds of measurement of this in indicator is that, although the official statistics about employment in the EU were not updated when we did the estimates for 2014, we nevertheless had data about the macroeconomic trends and about the main macroeconomic variables. We therefore knew how 2014 went.

For the year 2020 the uncertainty level about the general macroeconomic trend was clearly much higher. As already explained in the chapter about the forecasts (chapter 3), the forecasts for the 2020 consider three scenarios: a challenge, a baseline and a High Growth one.

The main assumptions adopted to forecast the data workers at 2020 are as follows:

- The estimates of data workers in 2014 represent the total number of employed workers. We have estimated the vacancies (unfilled positions) separately.
- The data worker forecast to 2020 for the three scenarios represents total potential demand, including potential vacancies (unfilled positions).
- Data is becoming a multipurpose production factor, adopted by most industries and companies.
- Using data in the production function improves the total factor productivity (TFP) of enterprises and industries.
- Data workers' growth is a function of data market growth, of data capital stock growth, and TPF growth.
- In the Baseline scenario and in the High Growth scenario, the TFP of the data industry and data users will significantly improve, compared to the general one.
- In the Challenge scenario, we assume that the use and organization of the resources/inputs of the data industry is less effective than it is in the Baseline scenario so that the TPF is similar to the TFP in the overall economy.
- In the next six years the data capital stock trend is expected to be similar to the ICT capital stock trend.
- In the Baseline scenario, we assume that the capital stock dedicated and accumulated for the data industry is higher than the ICT stock.
- In the Challenge scenario, we assume that the capital stock of the data industry has a similar trend to the ICT industry as a whole.
- In the High Growth scenario, the capital stock for the data industry has a trend increasingly more rapidly than in the Baseline scenario

The Occupational Mix of Data Workers

We also looked at data workers in terms of occupational mix: we selected the four most relevant occupational categories according to the latest International Standard Classification of Occupation (ISCO) as adopted by the International Labour Organization (ILO) and estimated the percentage of data workers who are actually active in each of the selected categories. The main results are shown in the figure below.

Figure 58 ISCO occupational mix of data workers, 2013
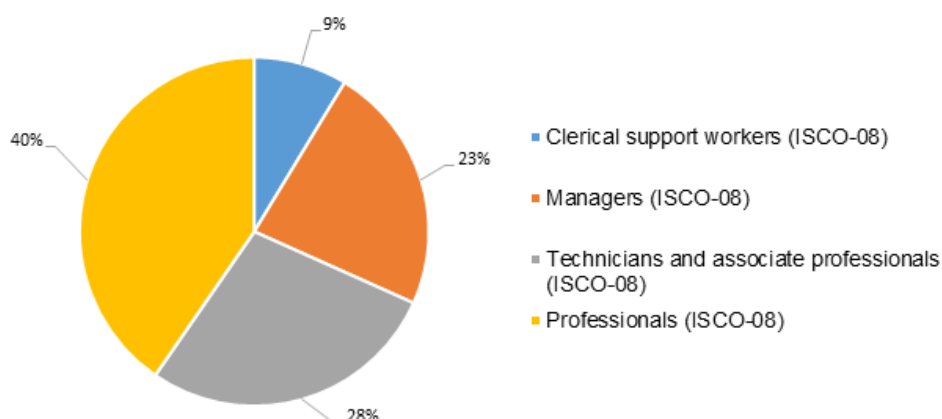
Source: European Data Market Monitoring Tool, IDC 2015

Overall, the occupational mix of the data workers shows that most of them are professionals, technicians, or professional associates. Based on a broad definition of data workers, and consistent with the survey conducted for this study, the data workers are distributed in all the selected ISCO occupations. Data workers as previously defined are in fact people with a medium-high professional level. There are also a significant number in the manager occupation category. Data workers in fact very much focus on data in order to drive their decision processes.

6.3. Measuring Data Companies

Statistical Classification of Potential Data Companies

Data companies are not classified in official statistics, so as a first step we have identified the main NACE sections where data companies based on our definition are likely to be found. The criteria used to define the perimeter of the data companies is as follows:

- We have included the NACE sections where specialized intermediaries and ICT enablers operate.
- In some NACE sections, although companies are not specialized intermediaries or ICT enablers, it may be that the companies (having a different core business) start new business units to collect and aggregate data products and services. For now, and for the sake of the study, we exclude these companies because they are currently a marginal part of the data industry.
- We have excluded companies collecting and implementing data products and services for their own use; we only consider data suppliers to be companies selling data products and services and therefore bringing in revenues.

Therefore we decided to focus only on the following 2 NACE rev2 sections to search for data companies. The share of data companies of the total was also measured on the total population of these 2 sections only:

- Section J: information and communication
- Section M: professional, scientific, and technical activities

SMART 2013/0063 European Data Market

The codes selected for both Section J and Section M are presented in the two tables below.

Table 52 Selection of codes from Section J, NACE rev2, where data companies may be classified

SECTION J — INFORMATION AND COMMUNICATION				
Division	Group	Class		
58			Publishing activities	Included
		58.12	Publishing of directories and mailing lists	Included
62			Computer programming, consultancy and related activities	Included
	62		Computer programming, consultancy and related activities	Included
		62.01	Computer programming activities	Included
		62.02	Computer consultancy activities	Included
		62.03	Computer facilities management activities	Included
		62.09	Other information technology and computer service activities	Included
63			Information service activities	Included
	63.1		Data processing, hosting and related activities; web portals	Included
		63.11	Data processing, hosting and related activities	Included
	63.9		Other information service activities	Included
		63.99	Other information service activities n.e.c.	Included

Table 53 Selection of codes, Section M NACE rev2, where data companies may be classified

SECTION M — PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES				
Division	Group	Class		
70			Activities of head offices; management consultancy activities	Included
	70.2		Management consultancy activities	Included
		70.22	Business and other management consultancy activities	Included
72			Scientific research and development	Included
	72.2		Research and experimental development on social sciences and humanities	Included
		72.2	Research and experimental development on social sciences and humanities	Included
73			Advertising and market research	Included
	73.1		Advertising	Included
	73.2		Market research and public opinion polling	Included
		73.2	Market research and public opinion polling	Included
74			Other professional, scientific and technical activities	Included
	74.9		Other professional, scientific and technical activities n.e.c.	Included
		74.9	Other professional, scientific and technical activities n.e.c.	Included

Statistical Classification of Potential Data Users

According to our research hypotheses, every company or organization is potentially a data user, so all NACE codes should be included. We have aggregated the main sectors to be able to develop a realistic sample and analysis based on the following table.

SMART 2013/0063 European Data Market

Table 54 Main industries and NACE codes where users may be classified

Industry segmentation	NACE rev2
	NACE section(s)
Mining, Manufacturing	B — C
Electricity, gas and steam, water supply, sewerage and waste management	D — E
Construction	F
Transport and storage	H
Information and communications	J
Finance	K
Public Administration And Defence; Compulsory Social Security	O
Education	P
Human health activities	Q
Wholesale and retail trade repair of motor vehicles and motorcycles, Accommodation and food services	G — I
Professional services, administrative and support services	L-M-N

Data Companies Forecast - Main Assumptions

Looking forward to 2020 and the evolution of the data market we have developed some assumptions on the development in the next years of the data industry and the data users, which represent important differentiators between our scenarios. They are briefly summarised here.

For data companies:

- The leading variable in our scenarios is the value of the data market, which is forecast on the basis of IDC research, the survey and the growth of overall ICT spending;
- The growth rate of the number of data companies in all scenarios is correlated with the data market growth, because they deliver the goods and services sold in the market (excluding imports);
- The growth rate of the number of data companies is lower than the growth rate of the data market, because we expect existing suppliers to increase their average revenues as well as new companies to enter the data industry;
- We expect medium-large companies over 250 employees to enter the data industry first. In 2014 the penetration rate (share of actual data companies of the number of potential data companies) of medium-large companies is more than double than the SMEs' one and it keeps growing in all scenarios.
- We also expect the absolute number of medium-large data companies to increase at a relatively slower rate than SMEs, since their number is already relatively high in

221

SMART 2013/0063 European Data Market

2014. In addition, the total number of medium-large companies is limited and it increases only when small companies grow over the 250 employees' threshold, so the pool of potential data companies does not grow as much. However, thanks to economies of scale the data industry is very attractive for medium-large companies so their number does increase in time.

- The structure of the ICT and professional services industries, which form the potential population of data companies, has a very high proportion of SMEs, and this does not change in the 3 scenarios. We expect the number of start-ups and spin-offs as well as small innovators to grow substantially in the High Growth scenario compared to the other 2 scenarios, which is consistent with the basic premise of a scenario of accelerated data market growth.
- The distribution of the data industry by Member State is correlated with the structure of the two industries from which data companies come in each country. We do not expect the landscape designed in 2014 in terms of the relative weight of each Member State in the data industry to change radically by 2020. For example, the U.K. will likely remain the country with the highest number of data companies in each scenario. This is because the main factors affecting the evolution of the scenarios are active across all of Europe, even if their influence is modulated by the local context.

Data Companies forecast model

- Taking data companies indicators in 2014 as the starting point, we estimated CAGR (compound average growth rates) for 2020/2014 for each scenario, modulating them by company size and MS. This was done taking into account the CAGR of the data market and the CAGR of total ICT spending by scenario.

The Member State calculation took into account the Member State clusters used in the survey to calculate the indicators for 2013-2014.

- The growth rates were used to calculate the absolute number of data companies by Member State and by size class.
- The share of forecast data companies on total is calculated on the basis of the number of total companies in 2014, since Eurostat does not provide forecasts of the total number of companies by industry. However, the structure of the European economy is not expected to change radically, even considering the balance between the creation and destruction of companies. Therefore, the forecast share indicator should be considered as a reliable proxy of the different scenarios outcomes.

Quality Control

- The results of the model (number of data companies) at the total EU28 level, by Member State and by size class were revised and cross-checked for consistency with the data workers' indicators, the data users indicators, the revenues indicators, the data market indicators. This was done:
  - By Member State verifying the coherence of the relative indicators of data market size and growth rate, data companies number and share of total J and M companies, the number of data workers and their share of employment, the number and share of data users.
  - By Member State verifying the coherence of the same indicators with the intensity of ICT spending on GDP and the growth trends of indicators with the growth trends of GDP and ICT spending.
  - By company size with the same approach across Europe.

Even if the growth rates of the model were built with correlations between indicators, still the absolute results once applied to the actual population of data companies per Member State had to be cross-checked for consistency and for a reality check.

In this phase for example we take into account extraordinary events such as the situation in Greece which is not yet reflected in the data employed for the model.

Data Users Forecast - Main assumptions

The following are the main assumptions used to forecast data users under the 3 scenarios.

- The leading variable in our scenarios is the value of the data market, which is forecast on the basis of IDC research, the survey and the growth of overall ICT spending;
- The growth rate of the number of data users in all scenarios is correlated with the data market growth, since data users represent the demand of the market.
- Our definition focuses on advanced data users. In the Challenge and Baseline scenarios we have used conservative assumptions about the growth rate of data users, assuming that they will be only the most innovative and competitive enterprises. In these scenarios the growth rate of data users is much lower than the growth rate of the data market. This is coherent with IDC's observed current rate of diffusion of Big Data technologies among EU companies (as analysed more in depth in the next paragraph).
- The leading assumption of the High Growth scenario is a change of pace in innovation adoption with a large increase of the number of data users. The data users' growth rate is almost as fast as the market growth; the growth rate of SMEs users is particularly fast.
- The average spending of data users is higher in the Baseline scenario than in the Challenge scenario; in the High Growth scenario is it also lower than in the Baseline scenario because the population of users is larger and the intensity of spending by the single user tends to be lower.

Data users Forecast Model to 2020

For the first round of measurement of data users we adopted the following procedure. The same methodology was used in the second and third (final) round of measurements of the same indicator.

- Taking data users' indicators in 2014 as the starting point, we estimated CAGR (compound average growth rates) for 2020/2014 for each scenario, modulating them by company size and MS. This was done taking into account the CAGR of the data market and the CAGR of total ICT spending by scenario.
- The MS calculation took into account the Member State clusters used in the survey to calculate the indicators for 2013-2014.
- The growth rate of the number of data users in the Baseline scenario is an extrapolation of the natural growth trajectory based on the supply-side trends estimated for 2013-2014;
- The share of forecast data users on total is calculated on the basis of the number of total companies in 2014, since Eurostat does not provide forecasts of the total number of companies by industry. However, the structure of the European economy is not expected to change radically, even considering the balance between the creation and destruction of companies. Therefore, the forecast share indicator should be considered as a reliable proxy of the different scenarios outcomes.
- The growth rates were used to calculate the absolute number of user companies by Member State and by size class. The resulting penetration rates (share of user companies on total EU enterprises) by scenario were calculated and cross-checked

with the number of user companies at the EU level and by Member State to perform a "reality check" on the growth rates.

- The penetration rates are based on the number of total EU companies at 2014, because Eurostat does not forecast the number of companies.

Quality control

- The results of the model (number of data companies) at the total EU28 level, by Member State and by size class were revised and cross-checked for consistency with the data workers indicators, the data companies indicators, the revenues indicators, the data market indicators. This was done:
  - o By MS verifying the coherence of the relative indicators of data market size and growth rate, data companies number and share of total J and M companies, the number of data workers and their share of employment, the number and share of data users.
  - o By MS verifying the coherence of the same indicators with the intensity of ICT spending on GDP and the growth trends of indicators with the growth trends of GDP and ICT spending.
  - o By company size with the same approach across Europe.

Even if the growth rates of the model were built with correlations between indicators, still the absolute results once applied to the actual population of data companies per Member State had to be cross-checked for consistency and for a reality check.

In this phase for example we take into account extraordinary events such as the situation in Greece which is not yet reflected in the data used for the model.

Measuring Data Revenues

The following are the main assumptions used to forecast data revenues under the three scenarios. They are:

- The leading variable in our scenarios is the value of the data market, which is forecast on the basis of IDC research, the survey and the growth of overall ICT spending;
- The leading variable for the calculation of revenues is the average revenues per data company at the total EU level, by Member State and by size class;
- In the Challenge scenario, we expect the average revenues of large companies over 500 employees to grow faster because they are stronger in a slow growth scenario; in the other scenarios the average revenues of SMEs and medium size companies (between 250 and 500 employees) grow faster than those of very large companies.

Data Revenues forecast model

For the first round of measurement of data revenues we adopted the following procedure. The same methodology was used in the second and third (final) round of measurements of the same indicator.

- Taking data revenues indicators in 2014 as the starting point, we estimated the absolute level and CAGR (compound average growth rates) 2020/2014 of the average revenues per data company for each scenario, modulating them by size class;

- Then we repeated the process estimating the average revenues per data company per Member State, taking into account data market and data companies growth rates per Member State;
- Then we calculate total revenues by scenario by multiplying average revenues per the number of total data companies per MS and by size class.
- The value of total revenues by scenario is then cross-checked with the value of the data market, because revenues are equal to the market value minus imports plus exports and we expect the balance of import/export to represent a small share of the total market value, for the reasons described in par.5.4.1.

Quality control

- The results of the model (average and total revenues) at the total EU28 level, by MS and by size class were revised and cross-checked for consistency with the data companies, data workers and data market indicators. This was done:
    o By MS verifying the coherence of the relative indicators of data market size and growth rate, the number of data workers and their share of employment, the number and penetration of data companies.
    o By MS verifying the coherence of the same indicators with the intensity of ICT spending on GDP and the growth trends of indicators with the growth trends of GDP and ICT spending.
    o By company size with the same approach across Europe.

Even if the number of data companies was a factor in calculating total revenues by MS, it is still important to review both data companies and total revenues indicators by MS as a "reality check" of results. This helps to find out and eliminate possible mistakes. In this phase for example we take into account extraordinary events such as the situation in Greece which is not yet reflected in the structural data used for the model.

Measuring Data Market Value

Estimate of the Data Market Value

IDC measured this market based on four key components:

1. Data
2. Software
3. Hardware
4. IT services

Data is acquired, acted upon, and sold within the data market. Corporate entities trade data internally in their organizations and between organizations. However, there is no reliable mechanism to measure this very specific trade and any estimates would be no better than guesswork, so for this model the value of this traded data is not included in the estimate of the value of the data market. Software associated with the data market is application software that falls under the business analytics product category or system management software. The business analytics tools included for analysis and delivery of data in the data market are listed in the table below, which shows the share of each tool used to estimate the size of the data market for the Member States.

Table 55 Business Analytics software markets included in the model for the data market size.

Share attributed to the data market		
BA Software — Competitive Market	2013	2018
Advanced Analytics Software (Standalone and Embedded)	5%	40%
Content Analysis Tools	1%	40%
Data Warehouse Generation	10%	40%
Data Warehouse Management	5%	40%
End-User Query, Reporting, and Analysis (Standalone and Embedded)	5%	40%
Spatial Information Analytics Tools	10%	60%
Workforce Analytics Applications	0%	5%

Source: IDC April 2015

In addition to application software is system management software — mostly used to manage hardware associated with the collection, storage, and analysis of data. This system software IT spend is modelled on a percentage of the hardware market associated with the data market, which includes storage, server, and networking equipment. This tie ratio corresponds to between 60% and 70% of hardware IT spend, consisting only of disk storage, servers, and network hardware.

Hardware associated with the data market comprises server and storage for collection and storage of data, with some networking equipment included. Within servers most data work will involve high-end and midrange servers, so only a small share of volume server IT spend is included in this market.

IT spend in the data market associated with IT services varies over the development of the market. In the developmental phase a large share of IT services spend will focus on training, education, and planning. In the later years this will transition to operations, maintenance, and support. IDC's model incorporates this change of emphasis on IT spend. During this transition the share IT services IT spend associated with the data market maintains a very consistent percentage of the data hardware market at approximately 90% of this data hardware total.

The 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.

ICT Perimeter Value

The total ICT spending value sourced from IDC's Black Book includes the following categories of products and services.

SMART 2013/0063 European Data Market

Table 56 ICT Perimeter Value

The following IT categories are included:		
Master Categories	Primary Markets	Secondary Markets
Hardware	Networking equipment	Enterprise Networks
		Service Provider Equipment
	Peripherals	Printers
		MFPs
		Copiers
		Monitors
	Storage	Tape
		Disk Systems
	Systems	High-end servers
		Midrange servers
		Volume servers
		PC Desktops
		PC Notebooks
		Media Tablets
	Mobile Phones	Smartphones
Packaged software	Packaged software	Appl. development and deployment
		Applications
		System infrastructure software
Services	Services	Training and education
		IT consulting
		Implementation
		Operations management
		Support services
Telecom Services	Telecom Services	Mobile Voice
		Mobile Data
		Fixed Voice
		Fixed Data
ICT Total		

Note: Feature phones are excluded

Source: IDC April 2014

SMART 2013/0063 European Data Market

Measuring the Data Economy

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 approach to this estimate is provided in annex, with the specific assumptions necessary for the estimate of each component.

To finalize our estimates of the impacts in 2016-2020, the study team has also conducted additional desk research on a series of different internal and external sources, and the updates are basically based on macroeconomic trends and on ICT specific trends. The main sources used are (citare le stesse dello scorso anno) : (p. 132 vecchio rapporto)

General assumptions

- 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 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.
- The very difficult economic mood of the last two years leads us to assume that, for the induced impacts, additional earnings are not going to be spent as in the past.

Qualitative Stories - Methodology and Selection Criteria

The approach to the development of the stories has followed a staged methodology revolving around four main phases, as shown in Figure 58 below.

SMART 2013/0063 European Data Market

Figure 59 Phases of development of the stories
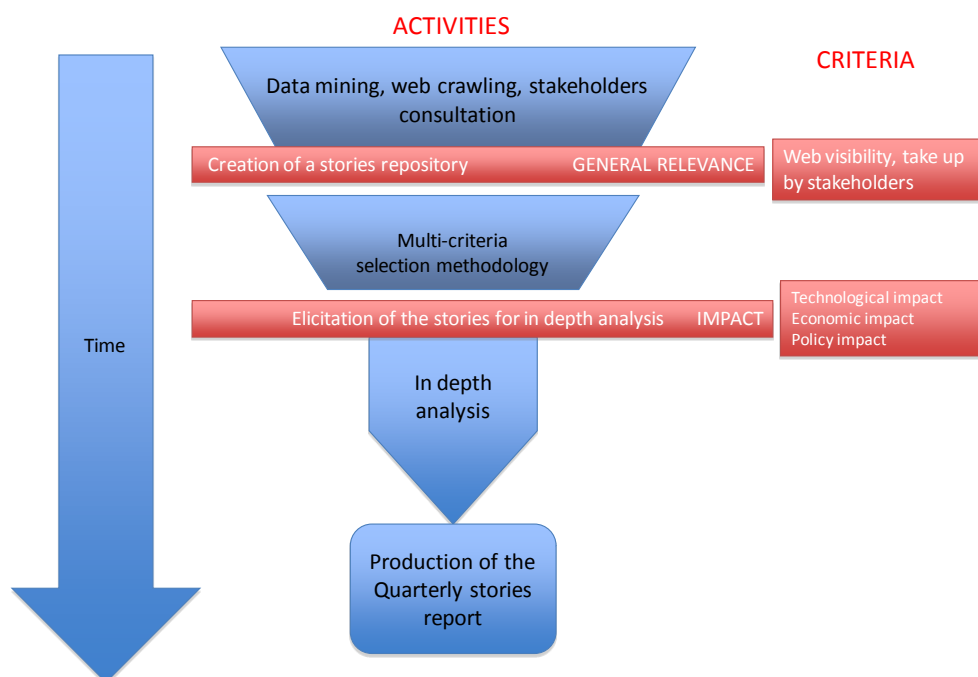
Source: IDC, 2015

Phase I: Data Mining and Stakeholder Consultation

In this first phase, the study team collected all applicable information, facts and figures, and potential case studies about the relevant aspects to be explored. A specific activity of data mining and desk research, across a number of different data sources and through the engagement of the stakeholders' community, has also been undertaken to identify a preliminary series of appropriate case studies to be further selected and used to develop the stories.

Phase II: Creation of a Case Repository

The selected cases have been used to populate a repository for the development of the stories; the repository has been used both as a means to initially submit the selected cases to the stakeholders' community, as well as a tool for classifying and prioritizing the cases.

Phase III: Development and Selection of Stories

The study team then selected — in accordance and coordination with the European Commission — a few key themes and developed a limited number of potential stories outlines, anticipating how they could be expanded into a story narrative. In parallel, the team has examined the facts and case studies collected and correlated them with these potential stories, by selecting the most compelling and relevant evidence of transformational impacts, ability to meet key Challenges, improved competitiveness, overcoming barriers to innovation, and so on. In doing so, the candidate stories have been analysed and prioritised on the basis on the specific criteria, such as:

- Availability of quantitative data

SMART 2013/0063 European Data Market

- Significance of impact
- Geographical balance
- Balanced coverage of the various aspects of the data value chain
- Relevance for the most recent policy developments

Finally, the original outline was modified to match the evidence and develop the final narrative of the story.

Phase IV: Story Production

For each selected story, the team carried out an in-depth analysis of the available documentation that was identified in phase 1 of the research. In-depth desk research, as well as qualitative interviews where possible were carried out to gather additional facts and figures.

Wherever possible, the team then drafted and developed the narrative of each story based on a common template whose main structure is as follows:

- Summary of the story: A short, standalone summary highlighting why the story is interesting for the European data market, the main findings, and the lessons learned.
- Introduction and general information: a table shows all the information necessary to map the stories within the data value chain.
- The story: description of the story, of the key themes and how they are illustrated.
- The cases: real examples of how the story unfolds, brief descriptions, facts and figures, and the technology implemented, the innovation aspects.
- The stakeholders and the business model: a description of the main stakeholders involved in the story and of the adopted business model.
- The impacts: actual and expected impacts identified in the story.
- Conclusions and implications for the European economy.

Measuring the Data Workers Skills Gap

Data Skills Model Structure

The model calculates separately demand and supply for six large Member States (France, Germany, Italy, Spain, Poland, and U.K.) and for EU28, while the rest of EU is calculated as a difference between EU and the six large Member States mentioned above. The complexity of the model and the lack of sufficient detailed data for all of the Member States prevents us from calculating the indicator for each Member State.

This model builds on the data workers model presented in chapter 4. The model was implemented for the first release in 2015. In March 2016 the model was implemented again updating the source data, for example the Eurostat data and revising the 2014 results. The same was carried out in October 2016 leveraging fresh Eurostat data and revising 2015 results.

The starting point is the stock of data workers employed in the EU in 2014/2015 plus the number of vacancies (unfilled positions). The number of vacancies for the year 2014 is estimated on the basis of the data user survey, where specific questions were asked to data user companies (see Annex for the questionnaire). The number of vacancies for the year 2015 has been extrapolated.

230

However, since perfect markets do not exist by nature, both unfilled demand and unemployment can and do exist at the same time, due to demand-supply mismatches (because some workers are not in the same geographic area as demand, or are not suitable for the specific position, or because of inefficiency in the placement of potential workers, and so on). Therefore, in every given moment there is a minimum level of "natural" unemployment, which is a combination of frictional and structural unemployment. This natural unemployment level decreases in periods of high economic growth and/or if the match demand-supply improves.

The data skills gap indicator is simply the difference between the total demand for data workers in the EU in the year 2020, under two scenarios, minus total supply. If demand is higher than supply, there is a data skills gap. If supply is higher than demand, there is over-supply and therefore unemployment. When the gap is high, there is a supply-side problem; when the gap is low, there may still be recruitment problems in specific countries and industries if the data skills market is not sufficiently flexible and mobile.

Figure 60 Structure of the Data Worker Skills Gap Model
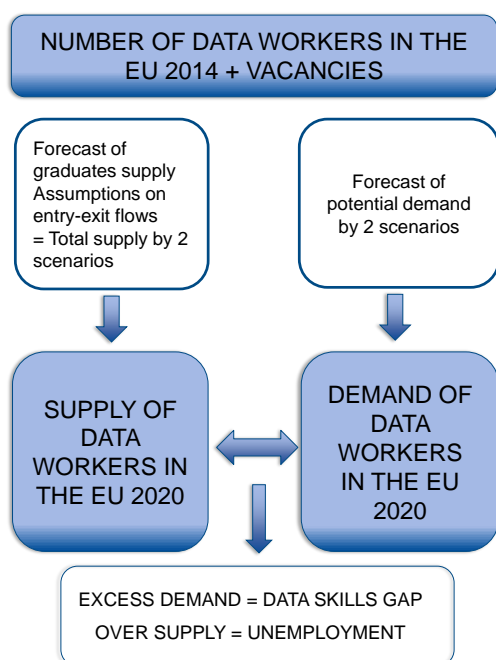
Source: European Data Market Monitoring Tool, IDC 2015

Calculation of Supply

The labour market is not a static entity, particularly in the case of an emerging market with a new typology of skills. As shown in Figure 58, there are both entry and exit flows which must be considered. To build the model, we developed general assumptions of entry and exit flows, plus specific forecast assumptions under the three scenarios, for each of the six Member States considered in this chapter and the EU as a whole.

SMART 2013/0063 European Data Market

Figure 61 Data Workers Supply-Side Dynamics
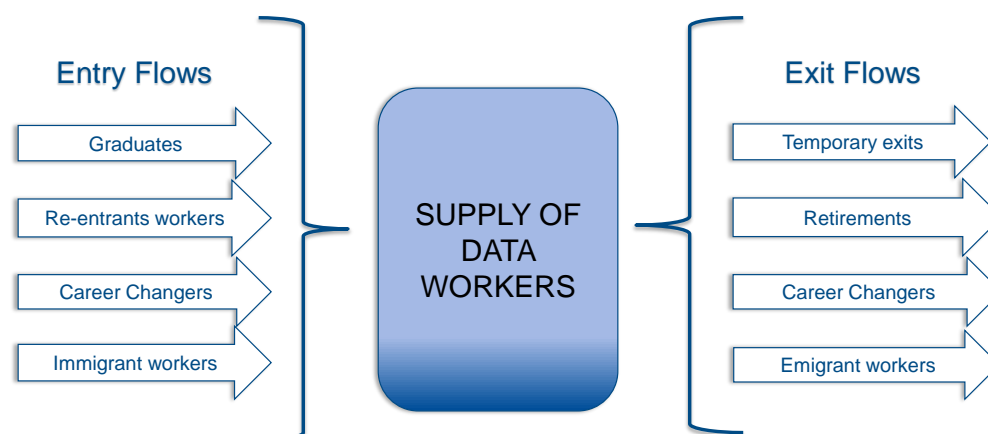
Source: European Data Market Monitoring Tool, IDC 2015

The main steps are summarised below with the underlying assumptions and reasoning provided on subsequent pages of this chapter:

- Calculation of total supply in the year 2014-2015-2016 based on:
  - o Total number of data workers employed
  - o Total number of vacancies
  - o Number of unemployed data workers (on the basis of actual unemployment rates by Member States)
- Forecast of additional supply for the period 2016-2020, under three main scenarios as follows:
  - o Estimates of entry flows:
    - ▪ Number of new graduates becoming data workers
    - ▪ Number of re-entrants from temporary exits (as a % on total)
    - ▪ Number of career changers (as a % on total)
  - o Estimates of exit flows:
    - ▪ Number of retirements (as a % on total)
    - ▪ Number of temporary exits (as a % on total)
    - ▪ Number of career changers (as a % on total)
  - o Estimate of the natural unemployment rate per Member States and total EU28
- Output: total number of data workers supply for six France, Germany, Italy, Poland, Spain and the U.K., the other Member States and the total EU28, under the three scenarios.

It should be underlined that the education system is relatively rigid: any change in enrolments will require a lag time of at least four years before it makes an impact on the number of graduates and therefore the labour market. The following general assumptions apply:

- The number of graduates will not alter drastically in the period to 2020 between the scenarios, due to the lead time between attracting students to relevant degrees and getting them to the level of graduation. However, initiatives such as the European Data Science Academy and graduate programmes initiated by the ICT vendor community will add additional supply to the market over the forecast period.

232

- Re-training and career changes represent the main sources of additional supply of data workers, with a much higher flexibility in time than the education system. Therefore, we assume lower training and inflows from other careers in the Challenge scenario compared to the Baseline scenario.
- There is a minimum level of unemployment called "natural", a combination of frictional and structural unemployment, which is almost impossible to eliminate or decrease substantially. Most sources estimate it between 2 and 5% of supply. In the case of our model, given that the current period is of high unemployment, we estimate a threshold around 5%.

Calculation of Graduates

The number of graduates with the appropriate skills is the most relevant source for data workers. The calculation of the number of graduates becoming data workers is based on the following steps:

- Identification of the main skills requirements for data workers (see Table 7 in this report);
- Selection of the main education courses providing data workers skills (see table below);
- Collection of data on the number of enrolments in the EU for the selected courses (Eurostat);
- Estimate of the % of graduates potentially becoming data workers;
- Forecast of the number of graduates expected in the period 2015-2020, by education. This was based on extrapolating the growth trends from 2008-2012 to 2016;
- Calculation of the number of graduates entering the labour market as data workers, on average four years after enrolment, in 2020 under 3 scenarios.

The profile of data workers includes a wide portfolio of skills, therefore the potential education path to this career includes many different options. We also expect the demand for "soft skills" of data workers (such as communication, team working, creativity) and business-related skills (specific knowledge of vertical markets) to increase in relevance over time.

Concerning the share of graduates becoming data workers, there is unfortunately almost no evidence, which lends a higher degree of uncertainty to the graduate element of the supply model. However, a sensitivity analysis of these assumptions indicate a relatively small impact on the total supply number in 2020 (see below). The approach was therefore to use as a reference quantitative indicator the % of data workers by ISCO category estimated by this study. The following assumptions were made:

- The main source of new data workers to the job market will be the Computing and Mathematics courses, since these will possess the core analytical and technical skills identified in Table 34 of this report. We assume that approximately one in four of these graduates will become data workers. If we instead assume that one in three of these graduates will become data workers, the total supply numbers in 2020 will increase only by 0.2%, indicating that strong swings in the proportions are needed to affect the supply situation materially.
- Application of data technologies is becoming increasingly important in the areas of life sciences and physical science. However, these fields of study are quite specialised so we would expect that the majority of the graduates would want to pursue careers within their specific area of expertise. However, some will pursue

233

careers that could be classified as data workers. We have assumed that one in twelve of graduates in life sciences and physical sciences will make this choice. If we instead assumed that this would be one in five, the total supply numbers in 2020 will increase by only 0.3%.

- Because of the need to extract business value from data-driven innovation and given their training in quantitative methodologies, we expect also that business and administration graduates may be attracted by data worker careers. However, many choose these degrees because of interest in e.g. strategy, marketing and organisational issues rather than data and numbers. At the same time, the career options for business and administration degrees are even wider than for science, mathematics and computing. Consequently, we assume that a lower proportion will dedicate their careers to working with data only – not using data as part of their other responsibilities. For the purpose of this study, we have assumed a 7% entry rate. If this entry rate would increase to, say, 10%, the supply numbers would increase by 0.4% by 2020.
- In the case of the other education courses identified, because of the specifics of the study areas, we expect that a small percentage of graduates may become data workers (ranging between 1 and 3%).
- These assumptions are in line with a study produced by Accenture[55], which found that 15% of available relevant talent chooses analytics careers. Similarly, a study by the U.K. Council for Economics and Business Research found that the ratio of relevant graduates who would choose analytics careers at about 17%.[56]

Table 57 Supply Estimates: Share of Graduates becoming Data Workers

Table	University Courses	Year of Graduation, 2020
Code	Discipline	Estimate % becoming Data Workers
EF4	Science, mathematics and computing	
EF48	Computing	23.0%
EF46	Mathematics and statistics	20.0%
EF42	Life science	8.0%
EF44	Physical science	8.0%
EF3	Social sciences, business and law	
EF34	Business and administration	7.0%
EF32	Journalism and information	3.5%
EF5	Engineering, manufacturing and construction	3.5%

---

[55] Accenture Institute for High Performance (2013) Crunch Time: How to overcome the looming global analytics talent mismatch.

[56] Council for Economics and Business Research (2013) Data equity – Ireland. Unlocking the Value of big data

EF31	Social and behavioral science	3.0%
EF14	Teacher training and education science	1.0%
EF2	Humanities and arts	1.0%
EF38	Law	1.0%
EF6	Agriculture and veterinary	1.0%
EF8	Services	1.0%
EF7	Health and Welfare	0.5%
UNK	Unknown	1.0%

Source: IDC estimate on Eurostat, 2015 and 2016

Main Forecast Assumptions

Regarding the forecast of data skills supply, and their related dynamics, the following assumptions apply:

Concerning the Baseline scenario:

- Unemployment of data workers is estimated at approximately 50% of the unemployment in each country, because of the positive growth trend of this market. On average, unemployment is around 5% that is very close to the "natural unemployment" threshold.
  - The lack of data scientists and big data analytics professionals is widely evident across Europe, as are the lack of certain programming and technology skills that are part of the data worker universe. For example, a report published in October 2014 highlighted that in the period from 2008 to 2013, there was a 212% increase in big data vacancies in the U.K. [57]. In addition, the then European Commission Vice President for Digital Agenda, Neelie Kroes, and Commissioner for Education, Culture, Multilingualism and Youth, Androulla Vassiliou were widely quoted for stating that there could be a shortage of up to 900,000 ICT skills in the EU by 2020. While many of these skills are outside the definition of data workers, there is a subset that is included – particularly around development and implementation of new data technologies, such as Hadoop. Some data workers' segments, such as clerical support workers, will experience high unemployment rates, but all in all in the assumption of halving the unemployment rate remains reasonable.
- Inflows from other careers: since data working is a growing field and there is emphasis on retraining / reskilling workers, assume that this adds an extra 5% to the supply a year on average (modulated by Member States depending on their characteristics, for example, it is higher for Germany because of the system's good capability of re-skilling). There is little available information on this particular aspect but one must assume that if segments of data workers are facing high demand, such as big data analysts, and that there is an emphasis on training

---

[57] Big Data Analytics: Assessment of Demand for Labour and Skills 2013–2020, The Tech Partnership and SAS, October 2014

more people in these specific skills, such as the initiatives supported by the European Data Science Academy and members of the IT vendor community, which are actively re-training and reskilling, there will be an increase in career changers. If this ratio was instead set at 10% a year, this would add around 5% to the supply.

- Outflows to other careers: we assume that fewer people leave for other careers than are joining estimated at –2% on average, modulated by Member State. This is based on the assumption that strong demand for data workers and retraining programmes will keep most data workers in their career path. There are no official statistics on this aspect. However, in case the outflow would increase to, say, -5%, the supply in 2020 would decrease by 2%.
- Retirement: based on current research and considering that the retirement age is being slowly increased around Europe, we assume a decrease of 1.5% per year on average. This is an elaboration of the results of the study *Monitoring e-Skills Demand and Supply in Europe*[58].
- Immigration-emigration: a large part of mobility is within Europe; the assumption is that immigration-emigration balance each other so the impact on the model is neutral.
- Temporary exits and re-entries: assumption that they balance each other out and the impact on the model is neutral.

Concerning the Challenge scenario:

- The inflows from other careers is lower than in the Baseline scenario, between 5-10% less, depending on the Member States;
- The outflows to other careers are on average double those in the Baseline scenario, because of lower demand and attractiveness of the data worker career;
- The retirement rate is slightly higher as data workers are encouraged to retire earlier (2% exit rate on average per year);
- Immigration-emigration and temporary exits-re-entries are estimated neutral, as in the Baseline scenario.

Concerning the High Growth scenario

- The inflows from other careers is higher than in the Baseline scenario
- The outflows to other careers re on average nil because of high demand and attractiveness of the data worker career
- The retirement rate is at the minimum because data workers are incentivised to retire as later as possible
- Immigration for professional reasons is incentivised and estimated positive from extra EU countries.

---

[58] Monitoring e-Skills Demand and Supply in Europe – Anticipating the Development of the Supply and Demand of e-Skills in Europe, 2010-2015, IDC and empirica, October 2009

Data workers' skills

Table 58 Description of the main data workers' skills

Skills	Description
Analytical core skills	Mathematics and statistics foundation Machine learning: elaboration of algorithms building models based on inputs to make predictions and decision making
Technical core skills	Programming based on various languages Database skills and use of distributed file system Use of open source tools and technologies
Domain and business knowledge	Knowledge of the industry or sector: data sources knowledge, understanding processes and trends behind data Understanding business goals and processes: knowing about the most relevant business issues and needs, organizational change and strategic developments.
Soft skills	Communication skills: ability to transform analytical insights into compelling stories and recommendations Team working: ability to work in team with people from different disciplines
Competencies	Analytical approach: able to focus on complex issues and questions Creativity: able to generate unexpected solutions to problems and exploring data from different angles.

Source: European Data Market Monitoring Tool, IDC 2015

Measuring Citizens' Reliance on the Data market – Methodological Approach

Citizens' indicators related to the use of ICT typically cover three dimensions: skills, adoption, and impact. Accordingly, we monitor:

- The capability of citizens to manage and understand data;
- The adoption by citizens of the listed solutions;
- The actual impact of using these solutions in terms of changing the behaviour of citizens.

The third variable is the most important as it deals with actual behavioural change driven by data. It can be considered a function of the other two variables, capability and adoption: decision making depends on having access to the necessary data (adoption) and on the capability to interpret that data.

At the same time, we need to consider which indicators are already available that could monitor this: the data for this indicator is at this stage not gathered through ad hoc data collection but by using existing datasets from different sources.

In the following table, we present an overview of the possible indicators and data sources to be used for the three dimensions, based on the criteria of data availability (at EU/MS level and across years) and data quality. Data quality includes the reliability (measurement is sound) and the validity (measurement covers the important issues).

SMART 2013/0063 European Data Market

The "ideal" indicators in terms of high validity are not available with the sufficient quality or geographical coverage, as the Table 79 in Annex shows. However, this gap could be solved in future editions either through Eurostat surveys or ad hoc studies.

With regard to capability, it is difficult to select one best indicator, since using data-driven apps requires a mix of thematic skills (such as health literacy), e-skills (to use the app) and quantitative skills (to interpret the data). At this stage, our choice is to use the DESI basic e-skills indicator (from Eurostat), because it's an official EU indicator; because it provides data in terms of the percentage of population; and because it includes advanced programming skills (which could be considered a proxy for quantitative skills).

With regard to adoption, the ideal indicator would be the percentage of citizens who use technology to track their lifestyle — as gathered in the PEW survey. In the absence of such data, and in the absence of robust data on apps, the closest proxy is the number of citizens owning a wearable devices, obtained by dividing the total number of wearable devices sold by the number of citizens in the country. The underlying assumption is that wearables are mainly a consumer device and that a person owns typically only one such device. Moreover, since data are only available from 2013 onwards, we assume that in 2014 there is no "replacement effect" so that all new purchases are from new customers.

With regard to impact, the ideal indicator would be the percentage of citizens who declare to have changed behaviour because of tracking data — as gathered in the PEW survey. There are no alternative indicators available.

This analysis shows that only two indicators fulfil all the criteria of high quality, relevance, availability for MS, and across several years:

- The DESI e-skills indicator by Eurostat. This will be the simple data provided by digital-agenda-data.eu.
- % of citizens owning a wearable computing devices. This will be the simple division of the number of devices sold in the last 2 years by the number of citizens in each country.

In conclusion, we propose to introduce a as the citizen's data indicator is the share of the population by country inclined to use data to inform personal behaviour decisions, using as a proxy the share of population with wearable devices and the appropriate skills".

The Survey: Methodology and Questionnaire

The secondary research and gap analysis conducted by IDC showed that existing data on data workers, the data market, and the data economy is very scarce. To offset this, the study team decided to collect primary fresh data through a field research effort. Both the supply side and the demand side of the data market were investigated through a field research survey of data companies and data users. The main objectives of the survey of data companies and data users were to:

- Gain a better understanding about the companies supplying data products and services, as well as the companies and organizations using and exploiting those data products and services
- Collect data about data company revenues and the employment of data companies and data users
- Gain a better understanding of performance and customers of both data companies and data users

238

SMART 2013/0063 European Data Market

Selection of the Survey Sample

IDC selected a sample of countries representative of the EU market. The parameters used to select the most representative countries and cluster them on the basis of their similarities were as follows:

Table 59 Indicators of selection of MS for the survey

Type of indicator	Indicators	Indicative sources
General	Population size	Eurostat
	GDP growth rate	Eurostat
	Geographic area (North, South, East, West)	Eurostat
ICT readiness and diffusion	IT spending on GDP (%)	IDC
	Fixed and Mobile Internet penetration	IDC
	Broadband connections penetration (incl. Urban/rural)	IDC, DAE
	Market concentration (leading operators market share)	NRAs-IDC
Culture and Technology	% of population having never used the Internet	Eurostat, DAE
	Digital literacy (competence) indicators	Eurostat, DAE

Such parameters were coupled with the need to keep the new survey within the limits of the time and budget constraints set out by the tender specifications to a sample of 8 countries:

1. The U.K.
2. Sweden
3. Czech Republic
4. France
5. Germany
6. Spain
7. Poland
8. Italy

The selected countries account for the vast majority of EU28 GDP, represent a good north-south and east-west balance, and provide a good balance of the different levels of IT sophistication. The EFTA countries are excluded from the survey.

Survey Sample

As displayed in the tables below, the proposed total sample of the survey consisted of 1,340 completed interviews. The actual sample size was composed of 1,437 completed interviews (a 7.2% increase on what was originally planned).

SMART 2013/0063 European Data Market

Table 60 Survey Sample

Sample Size: USERS			
	TARGET	ACHIEVED	METHOD
Czech Republic	100	110	CAWI
France	200	213	CAWI
Germany	200	216	CAWI
Italy	100	105	CAWI
Poland	100	100	CATI
Spain	100	116	CAWI
Sweden	100	109	CAWI
UK	200	215	CAWI
Total	1100	1184	

Table 61 Survey Sample's Size

Sample Size: Suppliers			
	TARGET	ACHIEVED	METHOD
Czech Republic	10	10	CATI
France	50	50	CATI
Germany	50	56	CATI
Italy	25	26	CATI
Poland	10	14	CATI
Spain	25	25	CATI
Sweden	20	22	CATI
UK	50	50	CATI
Total	240	253	CATI

Table 62 Survey Sample's Size Combined

Sample Size: COMBINED (USERS and SUPPLIERS)		
	TARGET	ACHIEVED
Czech Republic	110	120
France	250	263
Germany	250	272
Italy	125	131
Poland	110	114

SMART 2013/0063 European Data Market

Spain	125	141
Sweden	120	131
UK	250	265
Total	1340	1437

Interviews

The survey was conducted in the native language of each country. The total number of complete interviews provided a statistically representative sample in each country. The survey was based on a mix of CATI (computer-aided telephone interview) and CAWI (computer-aided web interview) techniques, whereby users were primarily interviewed through CAWI and suppliers through CATI.

Enterprise sizes are based on the number of personnel employed and were aggregated into two segments: fewer than 250 employees and more than 250 employees.

Respondents and Eligibility

Eligible respondents were, at each organization, the individuals best qualified to answer questions about the overall IT strategy and activities of the organization in the country. For medium/large organizations, the respondent was most likely the CIO, an IT director, or IT manager. For small organizations, it was more likely an IT manager or owner. A screening question determined eligibility.

Table 63 Survey sample segmentation by company size

Sample Size: USERS segmented by Company Size	
Company Sizes	ACHIEVED
1-9 employees	199
10-49 employees	250
50-249 employees	278
250-499 employees	171
500+ employees	286
Total	1184

Table 64 Survey sample segmentation by industry

Sample Size: Users segmented by industry	
Mining, Manufacturing	131
Utilities (electricity, gas, water, including sewerage waste and remediation activities)	45
Construction	103
Retail/ wholesale (including accommodation and food service	158

SMART 2013/0063 European Data Market

activities)	
Transportation and storage	90
Information and communications	118
Financial services (banking, insurance, other finance such as brokers, asset managers, stock exchanges, etc)	88
Professional services	164
Public administration and defence, compulsory social security	116
Education	109
Healthcare service providers	62
Total	1184

Data Source for Indicator 5

The data worker skills gap model used the answers to the following questions asked by the user survey, filtered by the number of respondents who were classified as data companies, according to the process explained in Chapter 4.1.3.

User Survey Questionnaire

Core questions — Section on Data Worker Level, Vacancies, and Skills

Q20. Thinking of the occupational matrix in your company could you indicate how many of these workers are considered data workers?

Please think of data as workers that collect, store, manage, and analyze data as their primary activity. They should be able to work with massive database and with emerging database technology to turn information into knowledge. Data workers include data analysts as well as data scientists but data entry clerks that perform very routine tasks related to data should not be considered.

		%
1.	Managers [Prog definition: managing directors, chief executives, other managers such as sales, marketing, ICT, etc.]	
2.	Professionals [Prog definition: [Prog definition: statisticians, researchers, software developers and analysts, database and network professionals, etc.]	
3.	Technicians and associate professionals [Prog definition: e.g., financial and mathematical associate professionals]	
4.	Clerical support workers [Prog definition: e.g., accounting & bookkeeping, statistical, finance, and insurance clerks]	

Q21. How many data worker related vacancies is your company looking to fill in 2015?

_____ [ABSOLUTE VALUE]

SMART 2013/0063 European Data Market

Don't know

Q21A. Do you expect your demand for data workers in 2015 to:

1. Increase significantly (>5%)
2. Increase somewhat (<5%)
3. Stay the same
4. Decrease somewhat (<5%)
5. Decrease significantly (>5%)
6. Don't know

Q22. Do you expect your demand for data workers in 2016 to:

1. Increase significantly (>5%)
2. Increase somewhat (<5%)
3. Stay the same
4. Decrease somewhat (<5%)
5. Decrease significantly (>5%)
6. Don't know

Q23. On a scale of 1 to 5 (1 = not at all, 5 = completely) rate the degree to which your organization already possesses each of the skills and qualifications for Big Data and analytics?

		1 Do not possess at all	2	3	4	5 Completely possess
1.	**Strategy development and project planning** and management (incl. defining KPIs, metrics, goals, resource requirements)					
2.	**IT hardware** (infrastructure management, maintenance, and tuning skills)					
3.	**Data analysis** using multi-dimensional analysis or OLAP, spreadsheets, visual discovery tools					
4.	Data analysis and exploration using advanced or **predictive analytics** methods such as statistical analysis, data mining, machine learning					
5.	Data or content (any type) collection, integration, preparation, management					
6.	Business intelligence and analytic **application development**, dashboard and structured report development					
7.	**Performance measurement** (e.g., monitoring of technology usage trends, post deployment decision quality and business outcomes evaluation)					

SMART 2013/0063 European Data Market

		1 Not success at all	2	3	4	5 Very success ful	Don't know
8.	**Governance** (e.g., technology or service evaluation, selection, and purchasing, external service provider or vendor management, information governance, security, and data privacy, master data management)						

Q24. How successful is your organization in hiring the following staff?

		1 Not success at all	2	3	4	5 Very success ful	Don't know
1.	Staff involved in strategy development and project planning and management						
2.	Business or program analysts						
3.	Data scientists, statisticians, data miners						
4.	Staff involved with data or content collection, integration, preparation, management						
5.	Staff involved in evaluating business and decision quality outcomes						
6.	Staff involved in governance processes						

Classification by Industry

The segmentation by industry sector used in the study is presented in the following table with the corresponding NACE rev.2 Codes.

Table 65 Industry Sectors Classification

Eurostat Name	NACE Rev 2 Code	Abbreviation for Tables
Construction	F	Construction
Education	P	Education
Electricity, gas and steam, water supply, sewerage and waste management	D-E	Utilities
Finance	K	Finance
Human health activities	Q	Healthcare

SMART 2013/0063 European Data Market

Information and communications	J	Information and communication
Mining, Manufacturing	B-C	Mining, Manufacturing
Professional services, administrative and support services	L-M-N	Professional services
Public Administration and Defence; Compulsory Social Security	O	Public Administration
Transport and storage	H	Transport
Wholesale and retail trade repair of motor vehicles and motorcycles, accommodation and food services	G — I	Wholesale / Retail

STATISTICAL ANNEX

Indicator 1.1 Number of Data Workers

Table 66 Number of data workers by Member State, 2014, 2015, 2016, 000s; % on Total Employment, 2016

Member State	2014 (000)	2015 (000)	2016 (00)	2016, % on total employment (*)
Austria	102	101	103	2.7%
Belgium	140	121	143	3.3%
Bulgaria	67	65	78	2.9%
Croatia	36	35	38	2.9%
Cyprus	9	9	9	2.8%
Czech Republic	122	129	133	2.8%
Denmark	70	70	76	3.1%
Estonia	20	18	20	3.6%
Finland	68	68	70	3.2%
France	678	705	739	3.1%
Germany	1,107	1,140	1,169	3.1%
Greece	66	55	66	2.2%
Hungary	100	103	104	2.9%
Ireland	56	57	63	3.7%
Italy	489	482	472	2.3%
Latvia	26	25	26	3.5%
Lithuania	42	38	43	4.2%
Luxembourg	11	12	13	6.5%
Malta	5	5	6	3.6%
Netherlands	253	264	262	3.6%
Poland	408	421	483	3.6%
Portugal	107	104	105	2.8%
Romania	130	122	126	2.0%
Slovakia	49	49	54	2.5%
Slovenia	28	27	27	3.4%

Spain	363	350	356	2.3%
Sweden	155	153	156	3.5%
EU27	4,707	4,730	4,941	2.9%
United Kingdom	1,111	1,275	1,220	4.3%
EU28	5,818	6,005	6,161	3.1%

Source: European Data Market Monitoring Tool, IDC 2016

Data Source: IDC estimates on Eurostat Labour Force Survey by occupation and NACE II industry code, 2015

Table 67 Intensity Share of Data Workers (average number of data workers per User Company) by Member State, 2014-2015-2016

Member State	2014 (units)	2015 (units)	N 2016 (units)
Austria	7.7	7.8	7.6
Belgium	11.1	11.3	11.1
Bulgaria	22.8	22.8	26.0
Croatia	22.6	22.4	23.8
Cyprus	6.0	6.1	5.5
Czech Republic	19.4	19.5	20.3
Denmark	9.0	9.2	9.6
Estonia	12.7	12.9	12.6
Finland	8.7	8.8	8.8
France	9.9	10.5	10.4
Germany	10.6	10.9	10.8
Greece	6.2	6.0	6.2
Hungary	23.5	23.4	23.9
Ireland	6.8	6.9	7.5
Italy	5.7	5.7	5.4
Latvia	24.0	24.8	24.0
Lithuania	23.5	24.4	23.1
Luxembourg	7.4	8.0	8.7
Malta	7.2	7.6	9.2
Netherlands	10.6	10.9	10.3

Poland	32.1	32.2	37.6
Portugal	5.9	5.9	5.8
Romania	23.6	22.2	23.1
Slovakia	19.1	19.0	20.9
Slovenia	24.2	23.3	22.6
Spain	5.1	5.1	4.9
Sweden	9.3	9.4	9.0
Total EU27	9.5	9.7	9.8
United Kingdom	7.5	7.7	7.9
Average EU28	9.1	9.2	9.3

Source: European Data Market Monitoring Tool, IDC 2016

Data Sources: IDC estimates on Eurostat Labour Force Survey by Occupation and NACE II Industry Code 2015

Table 68 Number and Employment share of data workers by industry, 2013–2014 (000s)

Industry	2014 (000)	2015 (000)	2016 (000)	2014, % on total employment	2015, % on total employment	2016, % on total employment
Construction	119	123	127	0.8%	0.8%	0.8%
Education	437	451	460	2.7%	2.7%	2.8%
Finance	582	601	618	9.0%	9.2%	9.4%
Health	458	472	485	1.9%	2.0%	2.0%
Information & Communication	641	662	682	10.2%	10.4%	10.7%
Mining, Manufacturing	683	704	718	2.0%	2.0%	2.1%
Professional services	1,188	1225	1261	5.4%	5.5%	5.6%
Public Administration	365	376	385	2.4%	2.5%	2.5%
Transport	181	187	191	1.6%	1.6%	1.7%
Utilities	88	91	95	2.8%	2.9%	3.0%
Retail and Wholesale	1,076	1110	1138	2.6%	2.7%	2.8%
Total	5,818	6,001	6,161	3.0%	3.1%	3.1%

Table 69 Forecast and Growth Rate of Data Workers by Member State, 000s, 2014-2020, three scenarios

Member State	2020 Challenge Scenario (000)	2020 Baseline Scenario (000)	2020 High Growth Scenario (000)	CAGR 2020/2016 Challenge Scenario (%)	CAGR 2020/2016 Baseline Scenario (%)	CAGR 2020/2016 High Growth Scenario (%)
Austria	106	107	160	0.7%	0.9%	11.7%
Belgium	154	208	332	1.8%	9.8%	23.4%
Bulgaria	83	99	114	1.6%	6.2%	9.9%
Croatia	41	52	58	1.6%	7.9%	11.0%
Cyprus	10	11	14	2.6%	6.2%	12.8%
Czech Republic	133	134	177	0.0%	0.1%	7.4%
Denmark	80	88	130	1.0%	3.7%	14.2%
Estonia	21	26	37	1.2%	6.7%	16.1%
Finland	72	73	81	0.7%	1.0%	3.8%
France	779	821	1216	1.3%	2.7%	13.3%
Germany	1241	1326	1647	1.5%	3.2%	8.9%
Greece	67	70	96	0.2%	1.4%	9.6%
Hungary	111	143	147	1.6%	8.3%	9.1%
Ireland	65	69	80	0.9%	2.1%	6.2%
Italy	489	649	1018	0.9%	8.3%	21.2%
Latvia	27	29	42	1.0%	2.6%	12.3%
Lithuania	44	44	69	0.9%	0.6%	12.7%
Luxembourg	21	24	32	12.9%	16.6%	24.8%
Malta	9	11	15	11.6%	16.1%	25.3%
Netherlands	290	477	517	2.6%	16.2%	18.5%
Poland	483	568	743	0.0%	4.1%	11.4%
Portugal	108	133	158	0.9%	6.2%	10.9%
Romania	126	132	144	0.1%	1.3%	3.5%
Slovakia	57	86	116	1.3%	12.2%	20.9%
Slovenia	29	41	46	1.6%	11.0%	14.3%

Spain	357	413	629	0.1%	3.8%	15.3%
Sweden	168	336	316	1.8%	21.1%	19.2%
EU27	5,171	6,169	8,134	1.1%	5.7%	13.3%
United Kingdom	1295	1644	2296	1.5%	7.7%	17.1%
EU28	6,466	7,812	10,431	1.2%	6.1%	14.1%

Source: European Data Market Monitoring Tool, IDC 2016

Data Sources: IDC estimates on Eurostat Labour Force Survey by Occupation and NACE II Industry Code 2015

Indicator 2.2 Data Companies

Table 70 Number and Share of Data Companies by Member State, 2015-2016

Member State	Number of Data companies, units 2015	Number of Data companies, units 2016	Share of data companies of total J and M sectors, % 2016	Total EU Companies 2016 Total J and M companies
Austria	4,250	4,300	14.9%	28,900
Belgium	2,300	2,350	11.5%	20,350
Bulgaria	1,100	1,150	5.5%	20,850
Croatia	650	650	5.6%	11,650
Cyprus	450	450	12.9%	3,500
Czech Republic	1,900	1,950	4.9%	39,700
Denmark	3,600	3,700	13.8%	26,750
Estonia	450	450	5.2%	8,600
Finland	2,750	2,850	11.9%	23,850
France	12,050	12,300	10.6%	116,500
Germany	24,950	25,500	14.4%	176,500
Greece	5,350	5,450	13.9%	39,350
Hungary	3,200	3,250	5.4%	60,450
Ireland	3,200	3,350	22.6%	14,800
Italy	18,400	18,500	15.0%	123,500
Latvia	500	500	5.3%	9,450

Lithuania	600	600	5.1%	11,775
Luxembourg	450	450	12.3%	3,675
Malta	200	200	12.1%	1,650
Netherlands	5,200	5,300	16.0%	33,100
Poland	5,300	5,650	9.5%	59,350
Portugal	3,950	3,950	11.5%	34,350
Romania	5,300	5,450	5.3%	102,500
Slovakia	2,100	2,150	5.1%	42,200
Slovenia	500	500	5.3%	9,400
Spain	15,150	15,300	10.3%	149,200
Sweden	8,050	8,100	13.6%	59,550
Total EU27	131,900	134,350	10.9%	1,231,400
United Kingdom	117,200	120,500	20.8%	578,400
Total EU28	249,100	254,850	14.1%	1,809,850

Source: European Data Market Monitoring Tool, IDC 2016

Indicator 2.3 Data Users

Table 71 Number and Share of Data Users by Member State, 2013–2014-2015

Member State	2015 Number	2016 Number	Growth Rate 2016/2015	2016 Share of data users of total EU companies	2016 Total Companies EU
Austria	13,350	13,550	1.5%	7.1%	190,300
Belgium	12,750	12,950	1.6%	8.5%	152,550
Bulgaria	2,950	3,000	1.7%	1.8%	167,050
Croatia	1,600	1,600	0.0%	1.8%	90,150
Cyprus	1,550	1,600	3.2%	5.4%	29,700
Czech Republic	6,400	6,550	2.3%	3.0%	220,800
Denmark	7,850	7,950	1.3%	6.4%	123,500
Estonia	1,600	1,600	0.0%	3.9%	41,450
Finland	7,850	7,950	1.3%	7.1%	111,750
France	69,750	71,300	2.2%	7.9%	900,550

Germany	105,950	108,200	2.1%	8.1%	1,340,700
Greece	10,700	10,650	-0.5%	1.8%	588,150
Hungary	4,300	4,350	1.2%	1.6%	277,750
Ireland	8,250	8,400	1.8%	9.7%	86,750
Italy	86,450	86,650	0.2%	7.2%	1,211,600
Latvia	1,100	1,100	0.0%	2.3%	48,300
Lithuania	1,800	1,850	2.8%	2.9%	64,000
Luxembourg	1,500	1,500	0.0%	9.0%	16,700
Malta	650	650	0.0%	7.6%	8600
Netherlands	24,500	25,400	3.7%	12.0%	211,350
Poland	12,800	12,850	0.4%	2.1%	621,300
Portugal	18,000	18,000	0.0%	6.6%	273,700
Romania	5,450	5,450	0.0%	1.0%	544,650
Slovakia	2,600	2,600	0.0%	1.7%	157,550
Slovenia	1,200	1,200	0.0%	2.0%	59,550
Spain	71,750	72,050	0.4%	6.3%	1,146,450
Sweden	17,000	17,450	2.6%	7.5%	232,850
Total EU27	499,650	506,400	1.4%	5.7%	8,917,750
United Kingdom	151,100	154,650	2.3%	11.0%	1,408,950
Total EU28	650,750	661,050	1.6%	6.4%	10,326,700

Source: European Data Market Monitoring Tool, IDC 2016

Table 72 Forecast of Data Companies Number by Member State, 2020, three scenarios

Member State	2020 Challenge Scenario	CAGR 2020/ 2016 Challenge Scenario	2020 Baseline Scenario	CAGR 2020/2016 Baseline Scenario	2020 High Growth Scenario	CAGR 2020/2016 High Growth Scenario
Austria	4,450	0.9%	5,150	4.6%	5,950	8.5%
Belgium	2,500	1.6%	2,900	5.4%	3,400	9.7%
Bulgaria	1,200	1.1%	1,400	5.0%	1,550	7.7%
Croatia	700	1.9%	800	5.3%	900	8.5%
Cyprus	450	0.0%	550	5.1%	600	7.5%

Czech Republic	2,000	0.6%	2,350	4.8%	2,700	8.5%
Denmark	3,850	1.0%	4,500	5.0%	5,200	8.9%
Estonia	450	0.0%	550	5.1%	650	9.6%
Finland	2,950	0.9%	3,400	4.5%	3,900	8.2%
France	12,800	1.0%	14,800	4.7%	17,150	8.7%
Germany	26,600	1.1%	30,850	4.9%	35,350	8.5%
Greece	5,550	0.5%	6,600	4.9%	7,450	8.1%
Hungary	3,400	1.1%	4,000	5.3%	4,550	8.8%
Ireland	3,450	0.7%	4,050	4.9%	4,600	8.3%
Italy	19,100	0.8%	22,600	5.1%	26,400	9.3%
Latvia	550	2.4%	650	6.8%	700	8.8%
Lithuania	650	2.0%	750	5.7%	800	7.5%
Luxembourg	500	2.7%	550	5.1%	650	9.6%
Malta	200	0.0%	250	5.7%	300	10.7%
Netherlands	5,600	1.4%	6,600	5.6%	7,500	9.1%
Poland	5,700	0.2%	6,850	4.9%	7,900	8.7%
Portugal	4,100	0.9%	4,800	5.0%	5,500	8.6%
Romania	5,600	0.7%	6,500	4.5%	7,550	8.5%
Slovakia	2,250	1.1%	2,650	5.4%	3,100	9.6%
Slovenia	550	2.4%	650	6.8%	700	8.8%
Spain	15,600	0.5%	18,450	4.8%	21,350	8.7%
Sweden	8,750	1.9%	10,350	6.3%	11,600	9.4%
Total EU27	139,500	0.9%	163,550	5.0%	188,000	8.8%
United Kingdom	125,750	1.1%	146,700	5.0%	171,050	9.2%
Total EU28	265,250	1.0%	310,250	5.0%	359,050	8.9%

Source: European Data Market Monitoring Tool, IDC 2016

Table 73 Share of Data Users on total EU companies by Member State, 2020, %, three scenarios

253

Member State	2020 Challenge Scenario/ 2016 total companies	CAGR 2020/2016 Challenge Scenario	2020 Baseline Scenario/ 2016 total companies	CAGR 2020/2016 Baseline Scenario	2020 High Growth Scenario/ 2016 total companies	CAGR 2020/2016 High Growth Scenario
Austria	7.1%	0.1%	7.8%	2.1%	10.9%	11.1%
Belgium	8.7%	0.5%	9.5%	2.9%	16.5%	18.1%
Bulgaria	1.8%	0.4%	2.0%	2.4%	2.5%	9.1%
Croatia	1.8%	0.8%	2.0%	3.0%	2.6%	9.5%
Cyprus	5.6%	0.8%	5.9%	2.3%	7.9%	10.1%
Czech Republic	2.9%	-0.2%	3.2%	2.0%	4.4%	10.6%
Denmark	6.5%	0.2%	7.0%	2.3%	10.5%	13.0%
Estonia	4.0%	0.8%	4.3%	3.0%	6.9%	15.5%
Finland	7.1%	0.0%	7.7%	2.0%	9.5%	7.6%
France	8.0%	0.3%	8.6%	2.2%	12.5%	12.1%
Germany	8.2%	0.4%	8.8%	2.3%	12.1%	10.6%
Greece	1.8%	-0.1%	2.0%	2.2%	2.3%	6.6%
Hungary	1.6%	0.3%	1.7%	2.5%	2.4%	11.4%
Ireland	9.7%	0.1%	10.5%	2.2%	13.3%	8.3%
Italy	7.2%	0.2%	7.9%	2.5%	13.5%	17.2%
Latvia	2.3%	0.0%	2.6%	3.2%	2.4%	1.1%
Lithuania	2.9%	0.0%	3.1%	2.0%	2.9%	0.0%
Luxembourg	9.3%	0.8%	9.9%	2.4%	15.0%	13.6%
Malta	7.6%	0.0%	8.1%	1.9%	12.2%	12.7%
Netherlands	12.4%	0.7%	13.6%	3.1%	21.9%	16.2%
Poland	2.1%	-0.2%	2.3%	2.3%	3.3%	12.0%
Portugal	6.6%	0.2%	7.2%	2.4%	9.5%	9.5%
Romania	1.0%	0.0%	1.1%	1.8%	1.4%	9.2%
Slovakia	1.7%	0.5%	1.8%	2.8%	3.1%	17.5%
Slovenia	2.0%	0.0%	2.3%	3.0%	3.0%	10.7%
Spain	6.3%	-0.1%	6.9%	2.2%	9.9%	12.1%
Sweden	7.8%	1.0%	8.6%	3.5%	14.1%	17.1%

Total EU27	5.7%	0.2%	6.2%	2.4%	9.2%	12.8%
United Kingdom	11.2%	0.4%	12.1%	2.5%	19.7%	15.8%
Total EU28	6.5%	0.3%	7.0%	2.4%	10.6%	13.5%

Source: European Data Market Monitoring Tool, IDC 2016

Indicator 3 Data Revenues

Table 74 Total and share of data revenues by Member State, 2015-2016, € Million

Member State	2015 € Million	2016 € Million	Growth Rate 2016/2015	Share on total revenues 2015
Austria	1,305	1,379	5.6%	4.3%
Belgium	1,461	1,477	1.1%	2.9%
Bulgaria	127	160	26.5%	3.0%
Croatia	85	83	-2.3%	2.4%
Cyprus	92	95	3.5%	9.9%
Czech Republic	342	313	-8.5%	2.1%
Denmark	1,404	1,609	14.6%	5.0%
Estonia	184	198	7.6%	12.0%
Finland	963	918	-4.6%	4.9%
France	7,338	7,915	7.9%	2.7%
Germany	11,976	13,367	11.6%	3.8%
Greece	455	464	2.1%	4.4%
Hungary	336	317	-5.7%	2.5%
Ireland	928	911	-1.9%	1.7%
Italy	4,600	4,736	2.9%	3.3%
Latvia	109	118	8.4%	5.0%
Lithuania	168	182	8.2%	9.7%
Luxembourg*	110	120	9.2%	0.0%
Malta*	46	50	9.0%	0.0%
Netherlands	3,146	3,424	8.8%	3.9%
Poland	1,378	1,801	30.7%	3.5%

Portugal	909	945	4.0%	6.1%
Romania	451	456	1.2%	3.3%
Slovakia	168	200	18.9%	2.6%
Slovenia	65	57	-12.9%	2.6%
Spain	3,333	3,491	4.7%	3.6%
Sweden	2,133	2,432	14.0%	4.0%
Total EU27	43,609	47,216	8.3%	3.4%
United Kingdom	12,423	14,603	17.5%	3.5%
Total EU28	56,033	61,819	10.3%	3.4%

Source: European Data Market Monitoring Tool, IDC 2016 - *Missing Eurostat update

Table 75 Forecast of Data Revenues by Member States, 2020, € Million, %, three scenarios

Member State	2020 Challenge Scenario € Million	CAGR 2020/2016 Challenge Scenario	2020 Baseline Scenario € Million	CAGR 2020/2016 Baseline Scenario	2020 High Growth Scenario € Million	CAGR 2020/2016 High Growth Scenario
Austria	1,513	2.4%	1,825	7.3%	2,456	15.5%
Belgium	1,759	4.5%	2,368	12.5%	3,333	22.6%
Bulgaria	211	7.1%	307	17.7%	366	23.0%
Croatia	119	9.6%	188	22.9%	193	23.7%
Cyprus	142	10.5%	188	18.4%	269	29.6%
Czech Republic	520	13.5%	650	20.1%	727	23.4%
Denmark	1,707	1.5%	2,024	5.9%	2,872	15.6%
Estonia	261	7.2%	349	15.3%	486	25.2%
Finland	1,389	10.9%	1,574	14.4%	1,929	20.4%
France	9,833	5.6%	11,888	10.7%	17,426	21.8%
Germany	16,819	5.9%	19,526	9.9%	28,506	20.8%
Greece	492	1.5%	607	6.9%	704	11.0%
Hungary	439	8.5%	553	15.0%	646	19.5%
Ireland	1,222	7.6%	1,536	14.0%	1,981	21.4%

Italy	5,778	5.1%	7,508	12.2%	11,753	25.5%
Latvia	152	6.6%	204	14.7%	241	19.7%
Lithuania	212	3.9%	279	11.3%	327	15.8%
Luxembourg	153	6.1%	186	11.5%	254	20.5%
Malta	56	2.8%	75	10.6%	102	19.4%
Netherlands	4,102	4.6%	5,339	11.7%	7,304	20.9%
Poland	1,844	0.6%	2,398	7.4%	3,140	14.9%
Portugal	946	0.0%	1,179	5.7%	1,581	13.7%
Romania	480	1.3%	643	9.0%	931	19.5%
Slovakia	225	3.0%	305	11.2%	406	19.4%
Slovenia	92	13.0%	126	22.1%	167	31.1%
Spain	4,022	3.6%	5,204	10.5%	6,482	16.7%
Sweden	3,146	6.6%	4,422	16.1%	4,765	18.3%
Total EU27	57,633	5.1%	71,453	10.9%	99,347	20.4%
United Kingdom	16,524	3.1%	20,421	8.7%	31,361	21.1%
Total EU28	74,158	4.7%	91,874	10.4%	130,708	20.6%

Source: European Data Market Monitoring Tool, IDC 2016

Indicator 4.1 Value of Data Market

Table 76 Indicator 4.1 Data Market Value in the EU28 by Industry, 2015–2016 and growth rate 2016/2015

Vertical Industry	2015 € M UPDATED	2016 € M	Growth Rate 2016/2015
Construction	254	275	8.2%
Education	1,442	1,562	8.3%
Financial services	10,760	11,816	9.8%
Healthcare	1,676	1,846	10.2%
Information & communications	5,231	5,865	12.1%
Mining, Manufacturing	11,736	12,814	9.2%
Professional services	7,855	8,490	8.1%

Public administration	3,011	3,323	10.4%
Retail and wholesale	5,810	6,319	8.8%
Transport and storage	2,464	2,690	9.2%
Utilities	2,240	2,466	10.1%
Home	1,871	2,071	10.7%
Total	54,351	59,539	9.5%

Source: European Data Market Monitoring Tool, IDC 2016

Table 77 Indicator 4.1. Data Market Value by MS, 2015-2016, € Million, and growth rate 2016/2015

Member State	2015 € M UPATED	2016 € M	Growth Rate 2016/2015
Austria	1,188	1,257	5.9%
Belgium	1,431	1,540	7.6%
Bulgaria	197	249	26.3%
Croatia	144	164	14.0%
Cyprus	115	120	4.1%
Czech Republic	527	612	16.2%
Denmark	1,123	1,283	14.2%
Estonia	178	198	11.0%
Finland	882	981	11.2%
France	6,907	7,427	7.5%
Germany	11,749	12,925	10.0%
Greece	398	430	7.8%
Hungary	346	397	14.7%
Ireland	877	1,042	18.8%
Italy	4,537	4,606	1.5%
Latvia	105	117	11.9%
Lithuania	160	181	12.8%
Luxembourg	104	114	9.7%
Malta	49	53	9.3%

Netherlands	3,162	3,395	7.4%
Poland	1,340	1,691	26.2%
Portugal	828	860	3.8%
Romania	510	584	14.4%
Slovakia	274	322	17.8%
Slovenia	106	113	7.4%
Spain	3,131	3,261	4.1%
Sweden	2,217	2,304	3.9%
Total EU27	42,586	46,226	8.5%
United Kingdom	11,765	13,313	13.2%
Total EU28	54,351	59,539	9.5%

Source: European Data Market Monitoring Tool, IDC 2016

259

Table 78 Forecast of Data Market by Member States, 2020, three scenarios

Member State	2020 Challenge Scenario	CAGR 2020/2016 Challenge Scenario	2020 Baseline Scenario	CAGR 2020/2016 Baseline Scenario	2020 High Growth Scenario	CAGR 2020/2016 High Growth Scenario
Austria	1,354	1.9%	1,508	4.6%	2,055	13.1%
Belgium	1,904	5.4%	2,340	11.0%	3,305	21.0%
Bulgaria	304	5.1%	344	8.5%	378	11.0%
Croatia	199	5.0%	227	8.5%	248	10.9%
Cyprus	166	8.4%	180	10.7%	188	11.9%
Czech Republic	588	-1.0%	722	4.2%	983	12.6%
Denmark	1,455	3.2%	1,637	6.3%	2,267	15.3%
Estonia	231	4.0%	285	9.6%	386	18.2%
Finland	1,063	2.0%	1,154	4.2%	1,397	9.2%
France	8,751	4.2%	9,134	5.3%	12,646	14.2%
Germany	15,537	4.7%	16,359	6.1%	20,613	12.4%
Greece	438	0.5%	540	5.9%	590	8.2%
Hungary	486	5.2%	558	8.9%	661	13.6%
Ireland	1,162	2.8%	1,270	5.1%	1,522	9.9%
Italy	5,128	2.7%	6,327	8.3%	9,600	20.2%
Latvia	133	3.1%	161	8.1%	168	9.4%

Lithuania	204	3.0%	249	8.3%	256	9.0%
Luxembourg	136	4.6%	153	7.7%	207	16.1%
Malta	71	7.6%	78	10.0%	93	14.9%
Netherlands	4,632	8.1%	5,609	13.4%	6,755	18.8%
Poland	1,631	-0.9%	2,148	6.2%	2,891	14.3%
Portugal	964	2.9%	1,118	6.8%	1,322	11.4%
Romania	591	0.3%	636	2.2%	888	11.1%
Slovakia	379	4.2%	488	10.9%	672	20.2%
Slovenia	138	5.0%	169	10.6%	184	12.8%
Spain	3,283	0.2%	4,045	5.5%	5,544	14.2%
Sweden	3,462	10.7%	4,450	17.9%	4,814	20.2%
Total EU27	54,390	4.1%	61,889	7.6%	80,632	14.9%
United Kingdom	16,017	4.7%	17,748	7.5%	26,189	18.4%
Total EU28	70,407	4.3%	79,637	7.5%	106,821	15.7%

Source: European Data Market Monitoring Tool, IDC 2016

Table 79 Share of Data Market on ICT Spending by MS, 2015-2016, %; 2020 Baseline - 2020 Challenge - 2020 High Growth

Member State	2015 UPDATED	2016	2020 Baseline Scenario	2020 Challenge Scenario	2020 High Growth Scenario
Austria	10.2%	10.8%	14.1%	12.9%	15.2%
Belgium	10.2%	10.9%	17.8%	16.4%	20.0%
Bulgaria	9.1%	11.6%	15.0%	14.8%	14.1%
Croatia	7.2%	8.1%	10.7%	10.6%	10.1%
Cyprus	15.7%	16.1%	24.3%	22.9%	20.9%
Czech Republic	7.0%	8.0%	8.4%	9.1%	10.7%
Denmark	8.2%	9.4%	12.7%	11.7%	14.0%
Estonia	23.2%	25.3%	31.1%	33.7%	39.5%
Finland	8.2%	9.1%	11.8%	10.5%	10.9%
France	7.8%	8.3%	11.7%	10.0%	12.0%
Germany	9.9%	10.8%	15.5%	13.3%	14.5%
Greece	8.5%	9.3%	12.0%	11.8%	11.3%
Hungary	7.0%	7.9%	10.6%	10.7%	11.0%
Ireland	12.8%	15.0%	19.2%	18.4%	19.1%
Italy	8.5%	8.6%	10.6%	11.4%	14.9%
Latvia	13.6%	14.7%	18.0%	19.1%	17.3%
Lithuania	14.0%	15.2%	18.6%	19.9%	17.8%
Luxembourg	4.9%	5.4%	7.5%	7.1%	8.3%
Malta	13.1%	14.2%	21.5%	20.4%	21.2%
Netherlands	9.5%	10.1%	15.4%	16.2%	16.9%
Poland	8.6%	10.7%	11.3%	13.2%	15.4%
Portugal	11.0%	11.5%	14.9%	15.1%	15.6%
Romania	10.4%	11.4%	12.6%	11.9%	14.3%
Slovakia	9.2%	10.8%	14.1%	16.0%	19.0%
Slovenia	7.0%	7.5%	10.3%	11.1%	10.5%
Spain	7.6%	7.9%	9.3%	10.0%	11.9%
Sweden	9.9%	10.2%	17.2%	19.2%	18.0%

Total EU27	9.0%	9.7%	13.3%	12.7%	14.3%
United Kingdom	8.1%	9.0%	12.5%	11.9%	15.0%
Total EU28	8.8%	9.5%	13.1%	12.5%	14.5%

Source: European Data Market Monitoring Tool, IDC 2016

Table 80 Share of Data Market on ICT Spending by Industry, 2015-2016, %; three scenarios

Industry	share on total ICT spending 2015	share on total ICT spending 2016	2020 Challenge Scenario	2020 Baseline Scenario	2020 High Growth Scenario
Construction	5.9%	7.3%	6.5%	8.8%	8.3%
Education	10.6%	10.9%	14.1%	18.5%	20.2%
Finance	10.3%	11.1%	12.7%	16.9%	16.2%
Human health activities	10.4%	11.4%	12.2%	16.8%	17.8%
Home	4.7%	5.2%	7.3%	9.3%	9.3%
Information and communications	18.0%	19.9%	22.5%	30.8%	29.4%
Mining, Manufacturing	6.7%	7.2%	8.4%	11.2%	10.6%
Professional services	8.5%	9.7%	10.2%	13.6%	13.8%
Public Administration	8.8%	9.5%	11.2%	14.7%	14.6%
Transport and storage	12.8%	13.4%	16.4%	22.1%	21.6%
Utilities	9.5%	10.1%	11.1%	15.3%	15.1%
Wholesale and retail	4.2%	4.7%	4.3%	5.8%	5.6%
Total EU28	8.8%	9.5%	11.1%	14.8%	14.5%

Source: European Data Market Monitoring Tool, IDC 2016

263

Indicators 4.2 and 4.3 Value of Data Economy and Incidence of the Data Economy on GDP

Table 81 Data Economy Value by MS, 2015-2016, €M, and total impacts as % of GDP

	Direct impacts		Backward Indirect impacts		Forward Indirect impacts		Induced impacts		Total impacts		Total impacts/GDP	
Member State	2015	2016	2015	2016	2015	2016	2015	2016	2015	2016	2015	2016
Austria	1,140	1,148	48	57	4,647	4,695	1,592	1,683	7,426	7,583	1.94 %	1.92 %
Belgium	1,360	1,457	71	88	4,042	4,123	1,918	2,062	7,391	7,731	1.60 %	1.62 %
Bulgaria	191	190	6	6	68	68	264	333	529	598	1.03 %	1.11 %
Croatia	139	117	5	4	72	74	193	219	409	414	0.80 %	0.79 %
Cyprus	111	108	5	5	31	32	154	161	301	306	1.49 %	1.50 %
Czech Republic	503	371	25	22	664	701	707	820	1,898	1,914	0.99 %	0.95 %
Denmark	1,073	1,149	50	65	3,424	3,648	1,506	1,718	6,052	6,580	2.01 %	2.14 %
Estonia	170	177	8	9	58	59	239	264	475	510	2.00 %	2.09 %

Finland	841	728	42	40	2,147	2,161	1,183	1,313	4,212	4,242	1.78%	1.75%
France	6,628	6,844	280	349	20,186	20,839	9,259	9,943	36,352	37,974	1.47%	1.51%
Germany	11,216	11,629	533	681	46,778	47,414	15,748	17,303	74,274	77,028	2.18%	2.19%
Greece	386	389	12	15	301	317	534	575	1,233	1,295	0.61%	0.64%
Hungary	333	277	13	13	169	175	464	531	979	996	0.78%	0.75%
Ireland	839	783	38	49	2,984	3,049	1,176	1,395	5,037	5,276	1.75%	1.68%
Italy	4,375	4,380	163	191	16,895	17,670	6,081	6,166	27,514	28,406	1.49%	1.52%
Latvia	101	105	4	6	38	39	141	157	284	307	1.00%	1.05%
Lithuania	154	162	6	8	86	88	215	242	461	501	1.07%	1.12%
Luxembourg	99	102	5	6	109	113	139	153	352	373	0.59%	0.59%
Malta	47	48	2	3	59	65	65	71	173	187	1.72%	1.75%
Netherlands	3,024	3,186	138	175	8,392	8,813	4,238	4,545	15,792	16,719	2.06%	2.14%
Poland	1,285	1,218	55	59	1,089	1,113	1,796	2,264	4,225	4,653	0.86%	0.89%

Portugal	806	817	23	28	796	827	1,110	1,151	2,735	2,823	1.33 %	1.34 %
Romania	495	437	15	17	292	305	684	781	1,486	1,540	0.80 %	0.77 %
Slovakia	262	282	11	16	103	106	367	431	743	835	0.82 %	0.89 %
Slovenia	101	83	5	4	83	86	141	152	330	325	0.74 %	0.71 %
Spain	3,020	3,001	111	128	13,006	13,476	4,197	4,366	20,335	20,971	1.67 %	1.67 %
Sweden	2,121	2,491	96	128	2,799	2,908	2,971	3,085	7,987	8,611	1.57 %	1.62 %
Total EU27	40,817	41,678	1,770	2,171	129,318	132,964	57,081	61,886	228,985	238,699	1.57 %	1.62 %
United Kingdom	11,270	11,831	494	609	29,115	31,027	15,769	17,823	56,649	61,290	1.94 %	2.06 %
Total EU28	52,087	53,509	2,264	2,780	158,433	163,991	72,850	79,709	285,633	299,989	1.73 %	1.76 %

Source: European Data Market Monitoring Tool, IDC 2016

Table 82 Data Economy Value by MS, Baseline Scenario

Member State	Direct impacts	Backward Indirect impacts	Forward Indirect impacts	Induced impacts	Total impacts	Total impacts/GDP
Austria	1,154	93	5,304	3,555	10,106	2.59%
Belgium	1,414	146	4,808	5,516	11,884	2.64%
Bulgaria	242	14	80	811	1,148	2.34%
Croatia	151	9	82	536	778	1.72%
Cyprus	136	11	50	424	621	3.57%
Czech Republic	510	56	774	1,701	3,041	1.58%
Denmark	1,202	106	3,942	3,858	9,107	3.01%
Estonia	189	18	75	672	952	4.04%
Finland	905	78	2,551	2,721	6,255	2.79%
France	7,429	554	22,784	24,518	55,286	2.31%
Germany	12,751	1,110	54,363	41,130	109,353	3.20%
Greece	380	24	348	1,273	2,025	1.11%
Hungary	403	32	199	1,316	1,950	1.46%
Ireland	928	99	3,393	2,993	7,412	2.03%

Member State	Direct impacts	Backward Indirect impacts	Forward Indirect impacts	Induced impacts	Total impacts	Total impacts/GDP
Italy	4,467	346	18,361	14,332	37,506	2.13%
Latvia	110	10	51	379	550	1.83%
Lithuania	169	15	123	586	892	1.90%
Luxembourg	111	10	109	361	592	0.88%
Malta	58	5	59	184	306	2.72%
Netherlands	3,487	341	9,479	13,222	26,530	3.65%
Poland	1,293	121	1,234	2,037	4,685	1.03%
Portugal	843	47	867	2,636	4,393	2.28%
Romania	501	30	356	1,499	2,387	1.13%
Slovakia	303	32	115	1,151	1,601	1.83%
Slovenia	105	10	97	399	611	1.50%
Spain	2,890	221	14,306	15,313	32,731	2.61%
Sweden	2,570	231	3,236	10,489	16,527	3.27%
Total EU27	44,701	3,769	147,147	153,611	349,228	2.56%
United Kingdom	12,992	1,666	31,922	34,449	81,030	2.80%
Total EU28	57,694	5,436	179,069	188,060	430,259	2.61%

Source: European Data Market Monitoring Tool, IDC 2016

Table 83 Data Economy Value by MS, Challenge Scenario

Member State	Direct impacts	Backward Indirect impacts	Forward Indirect impacts	Induced impacts	Total impacts	Total impacts/GDP
Austria	1,285	84	4,797	2,202	8,368	2.23%
Belgium	1,738	119	4,706	3,095	9,658	2.12%
Bulgaria	274	13	79	494	860	1.79%
Croatia	172	8	83	324	587	1.24%
Cyprus	148	10	41	269	468	2.58%
Czech Republic	626	46	635	957	2,263	1.19%
Denmark	1,351	94	3,723	2,366	7,534	2.69%
Estonia	233	14	68	375	691	2.96%
Finland	982	72	2,278	1,729	5,061	2.30%
France	7,754	531	22,358	15,102	45,745	1.98%
Germany	13,426	1,054	53,640	25,617	93,736	2.87%
Greece	469	20	289	711	1,489	0.78%
Hungary	463	28	202	791	1,483	1.13%
Ireland	1,014	90	3,195	1,889	6,188	1.84%

Italy	5,511	280	17,647	9,031	32,469	1.82%
Latvia	134	8	44	216	402	1.40%
Lithuania	206	12	100	332	649	1.52%
Luxembourg	125	9	110	221	466	0.70%
Malta	64	4	63	116	247	2.40%
Netherlands	4,223	282	10,747	7,532	22,783	2.95%
Poland	1,703	92	1,034	1,139	3,968	0.76%
Portugal	979	40	842	1,567	3,428	1.70%
Romania	539	28	295	961	1,824	0.98%
Slovakia	391	25	115	617	1,147	1.18%
Slovenia	129	8	98	224	459	1.05%
Spain	3,560	180	12,616	8,402	24,758	2.05%
Sweden	3,303	180	4,005	5,629	13,118	2.55%
Total EU27	50,802	3,331	143,812	91,906	289,851	2.18%
United Kingdom	14,396	1,504	33,241	22,334	71,474	2.64%
Total EU28	65,198	4,834	177,052	114,240	361,325	2.26%

Source: European Data Market Monitoring Tool, IDC 2016

Table 84 Data Economy Value by MS, High Growth Scenario

Member State	Direct impacts	Backward Indirect impacts	Forward Indirect impacts	Induced impacts	Total impacts	Total impacts/GDP
Austria	2,451	177	7,537	6,595	16,761	3.78%
Belgium	3,437	288	6,338	10,611	20,674	4.00%
Bulgaria	422	22	109	1,214	1,767	3.03%
Croatia	263	14	112	797	1,185	2.24%
Cyprus	217	16	49	605	886	4.48%
Czech Republic	1,193	107	1,209	3,156	5,665	2.59%
Denmark	2,620	205	5,667	7,276	15,768	4.67%
Estonia	441	33	106	1,238	1,819	6.39%
Finland	1,664	133	3,838	4,484	10,118	4.09%
France	15,029	1,075	32,568	40,595	89,266	3.39%
Germany	23,683	1,957	80,972	66,172	172,784	4.60%
Greece	717	37	423	1,893	3,071	1.48%
Hungary	768	53	311	2,122	3,254	2.07%
Ireland	1,701	165	5,058	4,886	11,811	2.82%

271

Italy	11,707	735	26,621	30,819	69,882	3.56%
Latvia	196	15	57	539	807	2.24%
Lithuania	296	21	127	821	1,265	2.26%
Luxembourg	237	19	158	666	1,079	1.41%
Malta	106	8	72	298	485	3.82%
Netherlands	7,119	576	14,108	21,684	43,487	5.21%
Poland	3,210	229	1,877	9,282	14,599	2.55%
Portugal	1,620	77	1,231	4,245	7,173	3.31%
Romania	1,054	59	567	2,851	4,532	1.87%
Slovakia	752	61	167	2,156	3,136	2.94%
Slovenia	195	15	129	589	928	1.89%
Spain	6,832	425	19,769	17,798	44,825	3.26%
Sweden	5,003	351	4,754	15,454	25,561	4.27%
Total EU27	92,936	6,873	213,934	258,844	572,588	3.76%
United Kingdom	29,741	3,442	49,542	84,071	166,795	5.11%
Total EU28	122,677	10,315	263,476	342,915	739,383	4.00%

Source: European Data Market Monitoring Tool, IDC 2016

Indicator 6.1 Citizens' Data Market

Table 85 Percentage of citizen owning a wearable device, 18 Member States, 2016

Member State	% 2015	% 2016
Austria	1,33%	2,86%
Belgium	1,32%	2,30%
Czech Republic	0,84%	1,38%
Denmark	3,02%	5,22%
Finland	2,77%	5,32%
France	4,21%	5,99%
Germany	2,80%	4,53%
Greece	0,12%	0,42%
Hungary	0,46%	0,75%
Ireland	3,07%	4,57%
Italy	0,65%	1,74%
Netherlands	2,07%	3,59%
Poland	0,40%	0,68%
Portugal	0,54%	1,48%
Romania	0,07%	0,24%
Spain	1,40%	2,70%
Sweden	2,40%	4,50%
United Kingdom	7,50%	10,15%

Table 86 Percentage of citizens taking decisions based on data, selected Member States, 2016

Member State	% 2015	% 2016
Austria	2,13%	4,45%
Belgium	1,84%	3,80%
Czech Republic	1,41%	2,40%
Denmark	3,50%	6,95%
Finland	3,33%	7,04%
France	6,28%	10,48%
Germany	3,96%	6,80%
Greece	0,37%	0,95%
Hungary	0,80%	1,51%
Ireland	5,04%	10,17%
Italy	1,95%	3,98%
Netherlands	2,42%	4,93%
Poland	1,04%	1,68%
Portugal	1,55%	3,08%
Romania	0,51%	0,89%
Spain	2,66%	4,96%
Sweden	2,83%	6,27%
United Kingdom	9,08%	15,02%