What is “big data?”
Who is using it? And why?

In the modern world, data is being abundantly generated, driven by the rise of user-generated content, the digitalisation of traditional industries and services, the adoption of machine sensors monitoring everything from tractors to jet engines, and the plummeting costs of computing power and (cloud) data storage.¹ To date, many business-case studies have focused on the successful data-driven business models of large US-based corporations like Facebook, Google, Twitter and the like.² But the real tide that data will shift will come not with the arrival of new whales in an already overcrowded sea, but with the flourishing of many fish, minnows and plankton throughout the economic ocean as a whole. Much more than is widely known, new companies are embracing “big data” and data analytics, spawning a whole new generation of startups, which are themselves helping to pioneer new business models and deliver unique...
experiences to customers. This new echelon of nimble startups provides a fascinating case study, offering important lessons on the role big data will play in the economy of tomorrow and revealing largely unheralded strength throughout a little known – but increasingly important – part of the European economy.

In this policy brief, we set out to learn about the kinds of businesses being started in this area, and to study the role and increasing relevance of data-driven business models among European startups. Based on extensive interviews with leading European data-business startups, and written in collaboration with the partners of the European Digital Forum, we look at best practice and the emerging business models in the data analytics field with a special focus on new companies. Who is doing what and where?

What sectors is data analytics most destined to change? Where are the opportunities? And what are the major pitfalls? In the final section, we will detail actions that need to be undertaken at the policymaking level to improve the overall performance of the data-driven startup ecosystem and create more opportunities for new Internet-based businesses in Europe.

Among the key findings:

1. Many young European companies are embracing data-driven business models, though, ironically, empirical data on the size and scope of data-driven businesses in Europe remains scarce.

2. The emergence of new companies and business models based on “data analytics” and “big data” opens enormous possibilities for transformation, including much change that will be immensely beneficial to consumers. The end result will be many new innovative, life-enhancing products available at low cost; better analytics for managing large social challenges at the macro level, such as climate change and public health; as well as immense private-sector success for companies quick enough to seize the initiative and provide the new goods and services to customers.

3. Data analytics will be particularly transformational in a host of sectors vital to the functioning of today’s economy: finance, logistics, manufacturing, new product development, retail, media, online services, marketing, sales, healthcare, energy, utilities, transport and more.

---

1 Businesses collect trillions of bytes of information on customer transactions, suppliers and internal operations. Social networking platforms such as Facebook and Twitter enable users to share over 1.3 billion pieces of information and content per day. Humans create 2.5 quintillion bytes of data every day, and 90% of the data in today’s world has been created in the last two years. The volume of data is doubling every three years. The total amount of data on file is expected to reach nearly 45 zettabytes by 2020. See IBM, Big Data at the Speed of Business (Armonk: IBM, 2014); A. T. Kearney, Big Data and the Creative Destruction of Today’s Business Models (Chicago: A. T. Kearney, 2012); and Diane McDonald and Ursula Kelly, Value and Benefits of Text Mining (Bristol: JISC, 2012).


3 While intuitively understood, startups are not legally defined (unlike small- and medium-sized enterprises). In the absence of this formal definition, available statistics is not comparable, and most national statistical authorities do not provide indicators and information on startups. Even more so, obtaining objective statistics specifically on data-driven startups, and/or conducting a large-scale survey, appears to be an overwhelming task.

4 This study was written by Sergey Filippov, associate director of the Lisbon Council, and based on extensive interviews with 20 European startups in the data analytics field. A list of the companies interviewed appears in the appendix on pages 30-31.
Thanks to their inherent agility and dynamism, startups in particular are benefiting from the opportunities that this technology engenders, and are exceedingly well placed to experiment with new data-driven business models.

Agile and disruptive startups can adopt a variety of business models across the data-value chain, ranging from collecting data and transforming it to information (data analytics) to the use of this information for the optimisation of business processes (business intelligence) and the provisioning of consulting services pertaining to all aspects of big data. Software development companies with strengths in this field are also thriving.

Startups can access different types of data from a multitude of sources. Most commonly, three types of data are available: 1) open data, 2) enterprise data, and 3) personal data. The terms and conditions for access to these types of data vary greatly, with personal data being the most heavily regulated and therefore difficult to use for new services and business offerings. Typically, the maximum value can be captured through the merger of various types of data from multiple datasets.

Access to talent is another urgent requirement for data-driven businesses. The necessary skills range from basic technical skills to business intelligence. The occupation of “data scientist” will emerge as a leading profession in the 21st century. These highly-trained professionals are in great demand, and their numbers are set to grow tremendously over the coming years. The vast majority of startups face a challenge in recruiting data science talent.

The US has more data-driven startups, but data-driven startups are increasing in number and taking on more importance in the European economy. Hurdles preventing European startups from emulating the successes of their US competitors are numerous, ranging from the “usual suspects” of limited access to capital and excessive bureaucracy to barriers specific in the digital domain such as the fragmentation of the digital single market and confusing data protection regulations.

Startups in general, and data-driven startups in particular, tend not to be well represented in the policymaking process, both at the European Union and national levels. The risk is that legislative proposals directly impacting data-driven startups are being debated in the absence of these key players. If passed, many proposed rules could potentially block the development of new business models, with a congruent downward trend in new jobs and growth. For instance, excessively rigid data protection rules and additional limitations on the use of even anonymised personal data could stop the development of many of these companies in the bud, devastating an entire class of promising European data-driven businesses. The voice of these players should be more strongly pronounced in public policy debates.

Data privacy issues, in particular, are a concern for startups – as they are for individual citizens. Much more can be done to clarify the regulations which surround privacy, and help startups navigate this maze. The regulation should be strong enough to ensure confidence, but not so rigid as to stop the delivery in Europe of exciting new analytics-based services and the development of data-driven business models.

‘The end result will be many new innovative, life-enhancing products available at low cost and better analytics for managing large social challenges.’
Against this backdrop, Europe must urgently take pro-active steps towards completion of the digital single market, closing the data skills gap, creating a smart regulatory environment, facilitating access to financing and allowing greater access to data within the existing framework for privacy protection.

**Big data and startups**

In order to gain a deeper insight in the European “big data” startup landscape, the European Digital Forum conducted 20 interviews with CEOs and/or co-founders of data-driven startups from several European countries, including Belgium, Bulgaria, Estonia, Finland, France, the Netherlands, Serbia, Spain and the United Kingdom. Combined with a detailed literature review, we approached data-driven startups through a variety of channels, including BigData-Startups, a new European platform. We found a healthy, vibrant ecosystem; some of these startups even have foreign offices already and are dynamically expanding overseas.

We formulated several guiding research questions:

- What novel business models are available to exploit the potential of big data?
- How do companies access data? What are the constraints and barriers to exploitation of the value of big data?
- How do data-driven companies access and manage data analytics skills?
- How does the current regulatory environment impact operations of data-driven startups?
- What are the main opportunities for and hurdles to data-powered growth?
- How do data-driven startups plan to scale up their operations?
- What public policy actions are needed to stimulate data-based innovation and underpin sustained data-driven economic growth?

---

5 For a full list of companies interviewed, see the appendix on pages 30-31.
6 BigData-Startups is building a comprehensive inventory of data-driven startups in Europe, which will become available at the beginning of 2015. Already, a few hundred startups are registered. Its website is www.bigdata-startups.com.
7 There are also two US-based startups – Parknav and BizEquity – in the group. Though based in the US, they are rapidly expanding in Germany and in the UK, respectively.
8 Special thanks to Henri Verdier, head of Etalab, the French government’s portal of open public data, and since 18 September 2014, France’s first chief data officer, for sharing his insights with the project team.

‘Many startups are developing innovative ways of extracting value and devising new business models in this fast-growing area.’
‘Already, US-based startups in the data field have a distinct competitive advantage over their European rivals.’

Smart Policy Principles

1. Define a regulatory environment that ensures adequate consumer protection but also empowers and enables new data-driven businesses

2. Build confidence in the use of consumer data for data-driven business models by helping to explain that not all data-driven businesses are violating consumer rights or abusing their data. Equating all data-driven business with exploitation and abuse of data is counterproductive and holds adverse effects on society

3. Create a smart regulatory environment for startups, making compliance easier

4. Create a “29th regime” for digital companies, establishing a single legal format for companies that operate across the EU exclusively on the Internet

5. Move quickly to adopt a single data protection regime for all 28 EU member states

6. Support the open data movement and make more anonymised public data available for use by third parties

7. Facilitate the formation and evolution of a coherent data ecosystem and harness collaboration between startups and large corporate players

8. Promote data portability and foster common standards and algorithms for data analytics

9. Facilitate access to financing for startups and scale ups

10. Invest in skills and encourage educational opportunities in the data-analytics field

For more, see the discussion of smart policy principles that begins on page 25.
Data-driven business models
Essentially, data analytics allows organisations to analyse the future rather than the past, using predictive or even prescriptive models.

“...we create an ‘intelligence abstraction layer’ above the data assets. We get an understanding of the ‘data’ and can therefore start asking the ‘what if and what now’ questions that support the case for investment.” (Richard Gascoigne, DTP SolutionPath)

“We are in a journey from descriptive to predictive to prescriptive analysis. This is a high level journey, and most companies have barely taken the first step.” (Richard Edwards, DataShaka)

“We propose to bring analytics to the data; we create an ‘intelligence abstraction layer’ above the data assets. We get an understanding of the ‘data’ and can therefore start asking the ‘what if and what now’ questions that support the case for investment.” (Richard Gascoigne, DTP SolutionPath)

“We are in a journey from descriptive to predictive to prescriptive analysis. This is a high level journey, and most companies have barely taken the first step.” (Richard Edwards, DataShaka)

We found that successful startups — such as Belgium’s Swan Insights and the UK’s DataShaka — were the ones able to make sense of the huge amounts of information coming at them, and to help others make sense of the data flows as well.

“We are in a journey from descriptive to predictive to prescriptive analysis. This is a high level journey, and most companies have barely taken the first step.” (Richard Edwards, DataShaka)

“Big data’ is first and foremost a problem, before being an opportunity. You have more data than you can handle. To turn this problem into opportunities, organisations must understand the potential of big data, and dare to take the first step. The good news is that you don’t need to invest huge piles of money to get the first benefits.” (Laurent Kinet, Swan Insights)

“We see a lot of companies waking up to the big data opportunities. In many companies, data management has been the exclusive domain of geeks. Now it has come more mainstream. Good to see chief marketing officers talking about big data; there is clearly desire to act this way, and many projects kick off, but it is still in an early phase.” (Richard Edwards, DataShaka)

But access to data is the key to all of this, and combining data from different databases has the most potential to offer the best insights.

“The prospects lie in the critical mass – the combination of multiple databases and datasets. The real power of big data can be leveraged when 10, 15, 20 large databases are combined. Massive predictive patterns will be generated with this data.” (Asparuh Koev, Transmetrics)

Perhaps not surprisingly, we found that many startups are developing innovative ways of extracting value and devising new business models in this fast-growing area, and that the data itself comes from a variety of sources. Swan Insights, for one, says it often accesses data from as many as six unique sources, with its value added being its ability to create one useful narrative out of an otherwise difficult-to-read message.

“The open data movement needs to accelerate in Europe.”

We process “1) corporate data – from our clients: CRM, ERP, supply chain, industrial data; 2) social data – social graph of people on social networks – who’s connected to whom, cross-networks; and all publications on social networks; 3) open data – the open data movement is growing fast throughout Europe; 4) public data – all data that is available, mainly on the Internet; 5) acquired data – private data that we buy or get through partnership, e.g. B2B data; and 6) network data – all data captured from specific systems and networks – installed beacon networks or Wi-Fi, etc.” (Laurent Kinet, Swan Insights)
The global open data movement is taking off.

data.gov.uk is a UK government project, launched in January 2010, to make available non-personal UK government data open. data.gov.uk contains almost 14,000 datasets from various departments; all data is non-personal. data.gov.uk intends to increase the use of linked data standards, to allow people to provide data in a way that allows for flexible and easy reuse. All data on data.gov.uk is available under a worldwide, royalty-free, perpetual, non-exclusive licence.

The Open Data Institute (ODI) (www.theodi.org) is a UK non-profit private company. The ODI is dedicated to catalysing the evolution of open data culture to create economic, environmental and social value. The ODI operates the programme supporting startups using open data. With hands-on support from the ODI, startups are becoming part of an energising open data ecosystem. The first cohort of startups have secured over £2 million [or around €2.5 million] in contracts and investments from government departments, domestic and international businesses and other clients. Startups in the second cohort focus on smart cities, 3D printed toys and energy efficiency.

Etalab, the open data platform of the French government, established in February 2011, is tasked with co-ordinating government administrations and providing support to public administrative institutions to facilitate the widest possible use of public-sector data. It offers almost 14,000 datasets from over 300 organisations (central government, local authorities, public institutions) on a single inter-ministerial portal (www.data.gouv.fr). Etalab provides raw data in readable formats facilitating its use and re-use. Recently, Etalab launched the DataConnexions competition to identify innovative uses of open data. To date, more than 100 French startups have been selected and awarded. Its fifth edition will be held in February 2015. Etalab features and promotes these best cases of open data re-use on its website.

The EU Open Data Portal (www.open-data.europa.eu), launched in December 2012, is the single point of access to a wide range of data held by EU institutions, agencies and other bodies. It allows to easily search, explore, link, download and re-use the data for commercial or non-commercial purposes, through a catalogue of common metadata. Around 40 EU institutions, bodies or departments (Eurostat, the European Environment Agency, the Joint Research Centre and other European Commission directorates-general and EU agencies) have made high-value datasets available, making a total of over 6,500. Semantic technologies offer new functionalities. The metadata catalogue can be searched via an interactive search engine and queries. In addition to giving access to datasets, the portal also is an easy entry point to a whole range of visualisation applications using EU data.
Types and categories of data

There are three broad categories of data available to startups: 1) Open data in the public domain (released by government and private institutions), 2) enterprise data, and 3) personal data. We found that each has characteristics that are unique in and of themselves. And each is treated legally and administratively in very different ways.

1. Open data in the public domain. This is all the information that public bodies, as well as private institutions, produce, collect or pay for, and release in public domain, with free access to it. Data from public bodies is characterised by its quality as it is properly validated. Examples are geographical information, statistics, weather data, data from publicly-funded research projects, digitised books from public libraries, banks publishing the location of their cash dispensers. If such data is available through an application programming interface (API), it can offer real time information. The economic benefits of open data are massive, both for government itself and for the private sector – it drives revenue, cuts costs, improves efficiency and creates high-skilled jobs. The direct impact of open data on the EU27 economy was estimated at €32 billion in 2010, with an estimated annual growth rate of 7%. Many companies are building successful businesses around open data.

Madiva Solutions, a Spain-based startup, collects open data in the public domain from various sources, generated both by government and individuals. It subsequently subjects this data to rigorous treatment and analysis.

“Today, we are basically reading government databases (residential cadastral registry), real estate advertising websites and social networking websites. All the information we read is free and public. Data collection is just one part of a larger process where, after choosing the type and amount of data we need to read and after collecting it, we design an in-house application in which we embed all the data.” (Fernando Alfaro, Madiva Solutions)

While there exist good examples of trendsetters, many countries still have a long way to go. In general, the open data movement needs to accelerate in Europe. For Transmetrics, a Bulgaria-based startup that tackles the complex problem of capacity utilisation in cargo transport industry, more open data would be absolutely instrumental to fuel its business operations:

“More open data should be placed in the public domain. In the US, an online database allows users to freely browse through customs declarations. It would be helpful to have such data available in the EU: import-export declarations by commodity segments, down to individual import declarations. Governments can release more infrastructural data, on traffic information and road conditions and similar. It would greatly help all of us understand transport better.” (Asparuh Koev, Transmetrics)

10 Our respondents from The Netherlands praise their national statistics office, Centraal Bureau voor de Statistiek (CBS), for the variety, richness and reliability of its open data. Equally, the British respondents are positive about access to public sector data in their country.
2. **Enterprise data** is massively generated by organisations thanks to the increasing digitalising of most business operations. These vast amounts of data are normally shared by users within an organisation; typically, this is internal data on procurement, sales, accounting and so forth. Companies in certain industries (telecom operators, financial and insurance institutions and similar) generate huge volumes of customer data too.

Most data-driven startups in our sample operate in the B2B market, using mostly enterprise data and, sometimes, complementing it with data from other sources. Offering data analytics for clients – usually large, multinational companies – is the most common and straightforward work approach. The client grants access to its enterprise data, and a startup runs all of the necessary data operations and provides corresponding advice.

“We’re a service provider and a consultancy company. We advise and analyse our customers’ data to help them find new ways for gathering meaningful data and using analytics to get insight out of it. We share best practices and recommend the tools we’ve find out to be the most suitable for the case at hand.” (Markku Alanko, Ivorio)

In many instances, startups perform data analytics at the client’s premises.

“We execute projects for companies behind closed corporate doors. To avoid any problems, we never take enterprise data outside and perform all operations inside the company walls.” (Jan-Kees Buenen, SynerScope)

Companies can also use enterprise data to offer products for external clients, i.e., not for the company in possession of the dataset. Mezuro, a Dutch startup measuring mobility, counts the number of people and vehicle movements in an area on the basis of anonymous static data from mobile networks operated by a global telecommunication company.

“The first step is data extraction. We created a unique process of extracting data from mobile telecommunication networks, taking care of all European privacy regulations. The second step is a transformation process. We tailor results towards the needs of different clients – local communities and cities, commercial companies, shopping malls, etc.” (Paul Kremer, Mezuro)

 Whereas enterprise data remains the main source of data for consulting projects, companies acknowledge the extra value to be derived from the combination of internal enterprise data with open data. And correspondingly, a higher premium can be charged. The more data from various sources is taken in analysis, the thinner the uncertainty in analysis is.

“The real challenge is the different nature of data that exists. If you bring different parts of data together, you build a multidimensional picture of your client. The vast majority of data we handle – up to 80% – is enterprise data, including sales data, data from third party systems. The other 20% is public data.” (Richard Edwards, DataShaka)
“We combine two main types of data – open public data and internal enterprise data. Our clients grant access to their internal data and we combine machine-generated data, such as website log data, with structured data, such as application data inside companies.”
(Rob Dielemans, GoDataDriven)

Big data creates immense opportunities for companies offering consulting services pertaining to leveraging big data in the most effective way. They provide advice on data management along the whole data value chain, with a particular focus on the business intelligence phase. Many of them influence corporate decisions driving effective organisational change.

“We help big companies optimise their business processes – leverage predictive mechanisms and develop smart specialisation. In the past, you needed a lot of domain knowledge to do it; now you have lot of data, but you need to think carefully how to use it.” (Koen Havlik, Algoritnica)

Successful managers say it is essential that this approach be demand-driven. Such consulting projects should start with the articulation of the client’s need. Rather than seeing technical solutions in isolation, companies need to take a global view on how big data can optimise their client’s business.

“Company management has some ideas about big data, but we can provide specific solutions. First we need to understand what the company really needs.” (Fabien Girardin, Near Future Laboratory)

“The key though is not data itself, but insights. It is too much of pushing technologies to people, without thinking of what actually helps people. Many startups focus on algorithms, without understanding their clients. They should offer demand-led, not supply-led solutions.” (Bruce Hellman, uMotif)

3. Personal data generally relates to any recorded information about an identifiable individual, including name, address, sex, age, financial details, marital status, education, criminal record or employment history. Due to issues of privacy, access to personal data tends to be restrictive; however, individuals may voluntarily share their personal data through the informed consent process, generally understood as a freely-given, specific and informed agreement between an individual and an organisation regarding how the data the individual shares can be used and treated. Under EU law, personal data can only be gathered, analysed or shared with third parties under strict conditions. Persons or organisations collecting and managing information must protect the personal data from being put to uses other than those for which consent was given (which can include subsequent data analysis or any other function other than the purpose for which consent was originally offered). Perhaps not surprisingly, many data-driven companies have been pro-active in this critical and sensitive domain, taking steps to reassure consumers and, potentially, to give themselves a competitive advantage. Criteo, a company founded in France in 2005, working with Internet retailers to serve personalised online display advertisements to consumers, has appointed a global privacy officer to ensure effective compliance and to help the company respond dynamically to consumer concerns in this area.12

12 Criteo, Criteo nomme Estelle Werth au poste de Global Privacy Officer, 24 June 2013.

‘More startups will need to find their niches and focus. Now it is time to specialise.’

10 • European Digital Forum
Other data-driven startups have also found ways to do business while complying with the existing – and future – rules in this area, though they often complain that the rules are cumbersome, hard to understand and subject to frequent change.

“Data regulations are a minefield for a small company. You need to pay a hefty amount of money to professional lawyers to solve it. Legislation is very fluid and quite ambiguous in many instances; apart from the law, there are working parties around privacy, and it is not very clear how this will proceed.” (Paul Kremer, Mezuro)

For CitizenMe, a personal identity management service, personal data is at the core of the business model. The company helps people take back control of their digital profiles on social networks, apps and all other online content. CitizenMe says it manages to deal with personal data successfully because it acts as an exchange platform, without appropriating the clients’ data.

“We pull information from various sources on the Internet, get data from users’ devices, look at GPS, IP, Wi-Fi data. We operate with personal data but don’t hold any of it.” (StJohn Deakins, CitizenMe)

uMotif, a startup in the e-health sector, pursues the same model of being an exchange platform with no ownership of personal data. Users record their personal health data on the company platform themselves.

“Users upload their personal data themselves, and data remains their property. We offer a nice interface to help people collect their medical data.” (Bruce Hellman, uMotif)

Clinical Graphics, a Delft University of Technology spin-off company, operates in the e-health sector too. It provides a motion simulation service based on computerised tomography – CT scans. Orthopaedic surgeons around the world upload medical scans of their patients to the company website to receive a 3D model and motion simulations. Patients grant their informed consent for the company to use their personal data. Nevertheless, the issue of privacy remains a sensitive topic. Although largely harmonised under the 1995 data protection directive, each EU member state has its own personal data protection act, most of them stemming from late 1990s or early 2000s.13

“As a company registered and based in The Netherlands, we fully abide by the Dutch personal data protection act. If I tell it to someone from a UK hospital, they don’t know if the data will be properly secured or not. I need to explain that Dutch law offers a very high degree of protection to the patient. With a harmonised European data protection act it should be very simple. the same standards should apply across Europe.” (Peter Krekel, Clinical Graphics)

CT scans used by Clinical Graphics are fully anonymised. Anonymisation of personal data, an approach employed by many companies, is often seen as a promising solution to privacy concerns. Though there are concerns that future innovation might render the anonymisation more and more easy to break.

---

13 For a full overview, see European Commission, Directive 95/46/EC on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data.
“Proper anonymisation of data is getting more and more difficult as the number of heterogeneous data sources rises. A push for protecting customers by introducing data anonymisation is bound to fail due to that same reason. Each anonymous source provides a little more information about the same individual. And the more anonymous sources that describe the same individual are available, the more trivial it becomes to de-anonymise.”
(Jan-Kees Buenen, SynerScope)

As paradoxical as it may sound, while privacy concerns prevent many businesses from emergence and growth, at the same time they represent a new business opportunity and stimulate entrepreneurs who aim to generate new business offerings for privacy-conscious users.

“Markets will solve it. If someone wants more privacy, a company will seize this opportunity and put new products, such as encrypted phones, on the market.” (Koen Havlik, Algoritmica)

“If I were an entrepreneur in Europe, I would address fears of privacy and set up a European analogue of DuckDuckGo, an Internet search engine that emphasises protecting searchers’ privacy and avoiding the “filter bubble” of personalised search results. If am a former Nokia engineer left without job now, I’d set out to design a secure mobile operating system, addressing the issues of privacy.” (Michael Carter, BizEquity)

Bringing all of this together, we find that the data value chain consists fundamentally of three main areas of activity: 1) data collection, leading to the emergence of big data; 2) data analytics, which are applied to raw data to make sense of it and create a knowledge base; 3) the knowledge base, which enables data-driven decisions (business intelligence) that realise value-added growth and wellbeing. We combined the three areas of activity (data collection, data analytics, business intelligence) with the three types of data (open, enterprise and personal) to create a multidimensional model of business opportunities along the data value chain. See the table on page 13.

Working with several types of data along the data value chain may be necessary to deliver a high-quality product or service. It may also be a smart idea to specialise in a certain data operation to achieve internal efficiency.

“More startups will need to find their niches and focus. At the dawn of the big data age, many startups were founded, now it is time for them to specialise.” (Koen Havlik, Algoritmica)

But how are data-driven startups monetising the services they provide? We will turn to that question in the next section.

**Revenue models**

Generation of sustained revenue streams is at the core of any business model.14 The most common type of revenue model for data-driven companies, operating both in the B2C (consumer) and B2B (enterprise) markets, is subscription, where a customer must pay a subscription price (monthly/ quarterly/ annually) to have access to the product or service. A derivative model is a pay-as-you-go subscription, where a customer subscribes to purchase a product incidentally. Licensing – the granting of permission to use intellectual property rights under defined conditions – is a classic revenue model for software developers. In the B2B market, companies charge their clients for

---

14 The revenue model is the system design by which a business monetises its products and services.
Consultancy projects. Normally, consultancy projects are based on a lump sum (fixed amount for an entire project) or reimbursable (per hour consulting fees) contracts. The majority of startups in our study employ at least one of these revenue models.

“As many B2B companies, we have two main revenue streams: 1) a subscription model with a monthly fee for recurring reports, and 2) a project price for special (consulting) projects.” (Paul Kremer, Mezuro)

“We are in the enterprise market. We apply a hybrid model of consulting and software licensing.” (Jan-Kees Buenen, SynerScope)

“Working on a project basis is a good way to find out where demand is. We charge our clients per hour, or lump sum per project.” (Koen Havlik, Algoritmica)
As in any other type of business, data-driven companies do not need to rely exclusively on one revenue model; they can combine them to optimise revenue streams. Madiva Solutions, a Spain-based B2B, software as a service (SaaS) company, has four main types of data-powered revenue:

‘Monthly or quarterly subscription fees charged for the usage of our application or services; the actual charge is based on the number of licences granted; 2) non-recurrent services where we charge the client a fee per amount of data processed; 3) a periodical fee as in the first mode, where at the end of the year we adjust the revenue based on criteria like the level of commercial success achieved (conversion rates) and/or the level of usage (number of loggings to our application); 4) consultancy/collaboration agreements. We receive an annual ’consultancy’ fee on top of which we might make additional revenues based on specific projects accomplished.’

(Fernando Alfaro, Madiva Solutions)

If a company operates in the B2C market, next to subscription, it can also use an advertising-funded (advertiser-supported) revenue model. Advertising is the primary source of revenue – services are offered largely for free, but the user has to “consume” advertisement in exchange. Companies offer services in exchange for personal data, on the basis of which advertisement companies target their potential clients. Facebook, the most successful online social networking service, is a good example of an advertising-funded business.

Another commonly used pricing strategy in the B2C market is freemium. A product or service, such as software, media, games or web services, is provided free of charge, but a premium is charged for proprietary features, functionality or virtual goods. Such revenue model is particularly popular among mobile app developers. It can also span to the B2B market. Parknav, a startup launched in the US and expanding rapidly in Europe, offers its street parking information app for individuals free of charge. It is meant for individual drivers who occasionally face difficult parking situations, and limits the number of parking searches to 50 per month. Parknav’s business package is tailored to companies with large fleets, and is offered for $49.95 [around €40.00] per year.

CitizenMe, a UK-based startup, allows users to create their own digital portfolios of personal information. The company charges £10.00 [around €13.00] per year per user for offering personal identity management services (i.e. a classic subscription model). If users wish to share their personal data with advertisers, an advertising-funded model of the B2C2B market kicks-in. The service itself becomes free, and users are even rewarded monetarily (in cash, vouchers, air miles, etc.). Cash payments from advertisement companies may reach up to £10.00 [around €13.00] a month; and CitizenMe charges only 10% in its transaction fee. Why would advertisers pay money to people? In return, they receive very accurate, rich, up-to-date personal information, including explicit validated future purchase intent. Another example of a startup using the B2C2B model is uMotif, a UK-based health information platform. It offers its services for free to end users (who create their e-health profiles on the uMotif platform), but sells licenses to organisations that need to access these profiles, with users’ consent, such as insurance companies, the UK National Health Service and hospitals.
Getting the right skills

Any business model is as good as people who can implement it. Making big data business models work requires new combinations of skills. This includes database engineering and architecture skills to develop and maintain a big data infrastructure able to process large volumes of data rapidly and reliably. This requires expertise with new frameworks for data storage and manipulation (Hadoop, Cassandra, Hive, Pig, NoSQL), data analysis skills to access and clean data, extract insights from it and, in some businesses, develop software that transforms those insights into data-driven products and services. This involves a combination of analytical skills (including statistics and machine learning), coding and domain knowledge to be able to ask relevant questions, and understand how data fits within the business.15

In the late 2000s, Silicon Valley companies working with big data started referring to workers with data analytics skills as “data scientists.”16 A report by Nesta on data scientists refers to them as “unicorns,” something that is hard to find.17 Available empirical evidence supports this metaphor, and indicates significant big data skills shortages. Up to 80% of the new data scientist jobs created in the US between 2010 and 2011 were not filled.18 McKinsey Global Institute predicts that this situation is only going to get worse – the US will have a shortfall of 190,000 “deep data experts” by 2018.19 The situation in Europe is similar. Some 62% of the European managers surveyed in 2013 believe there are skills shortages around big data analytics.20 Research by e-Skills UK, the UK’s sector skills council for the ICT industry, shows that 57% of UK companies recruiting people with “big data skills” suffer difficulties finding the right talent.21 Our respondents corroborate these findings and acknowledge the difficulty of finding the data talent.

“The right mix of skills includes coding in combination with business acumen.”

In Model Workers: How Leading Companies Are Recruiting and Managing Data Talent, a recent report, researchers at Nesta examine how UK businesses are securing and managing talent to grow in a big data world.22 The study, based on in-depth interviews with UK businesses and educators at the forefront of the data revolution, and produced in collaboration with Creative Skillset and the Royal Statistical Society, considers how companies should recruit and manage their data talent.

15 See also Juan Mateos-Garcia and Andrew Whitby, Chasing Unicorns: Three Questions about Data Scientists (London: Nesta, 2013).
16 This title was coined by D.J. Patil and Jeff Hammerbacher in 2008, at that time the leaders, respectively, of data and analytics efforts at LinkedIn and Facebook. Thomas H. Davenport and D.J. Patil, “Data Scientist: The Sexiest Job of the 21st Century,” Harvard Business Review (October, 2012).
17 Mateos-Garcia and Whitby, op. cit.
18 Harlan D. Harris, Sean P. Murphy and Mark Vaisman, Analyzing The Analysers (Sebastopol: O’Reilly, 2013).
20 Teradata, Data Science in Europe – Skills Gap or Emerging Talent?  (Dayton: Teradata, 2013).
Of the 45 data-driven UK companies interviewed by Nesta in 2014, four in five reported difficulties recruiting analytical data talent. They provided three reasons for this:

1. Data talent in the market lacks the right level of skills and experience. Seasoned data analysts are very expensive, and junior people require extensive training because they do not have sufficient technical skills and hands-on experience working with data.

2. Many companies lack the capacity to recruit data talent effectively. There are hurdles in communicating data candidate specifications to HR managers and recruitment agencies, assessing the skills of candidates, and understanding the business value they can create. These challenges, which are heightened by lack of professional standards among data analysts, are particularly severe for companies that are new entrants in the big data arena.

3. Data talent lacks the right mix of skills. Good analysts often cannot code, and good coders often cannot analyse. Data analysts with the commercial instincts to produce business impacts are very rare. Some companies blame this on the existence of "silos" between data science academic disciplines (statistics, computing and business studies) in universities. This is consistent with the findings of Teradata’s research, where 58% of respondents report that big data candidates with the right mix of skills are hard to find.23

Going even further, the right mix of skills includes not only coding in combination with business acumen, the talent needs to possess plentiful soft skills.

> “It is indeed very difficult to find people with proper skills. We don’t need only computer skills, or analytical skills, or their combination; we also need soft skills – creativity, understanding of what society is, and what people are, understanding of the relationship between technology and society and its evolution.” (Fabien Girardin, Near Future Laboratory)

Unsurprisingly, rather than searching for “unicorns,” many data startups advocate for balanced multi-disciplinary teams consisting of people with complimentary, mutually reinforcing competences.

> “There are two schools of thought: 1) an individual “super-human” possessing all data and analytical skills – these people hardly exist, and 2) a team-based approach. We have database people, data scientists, statisticians and business analysts, all work as a team. You may find specialists in each of these areas and you can grow faster and easier this way.” (Asparuh Koev, Transmetrics)

Where can data scientists be found? What is the right education level? While European educational institutions fail systematically to deliver enough data scientists, there is no alternative, and they remain the primary source of data talent – if not graduates, then PhD candidates and research staff. As respondents assert, data analytics and business intelligence require people with advanced university degrees, able to think out of the box.

23 Teradata, op cit.
“We work with PhD-level academic experts having worked in fundamental research departments of targeted universities – Université Catholique de Louvain, M.I.T., Northeastern University, University of Ghent.” (Laurent Kinet, Swan Insights)

“We can tap into the academic world as we have a lot of good relations there. We activate scientific researchers for our projects.” (Fabien Girardin, Near Future Laboratory)

An alternative view is that it is not people that should master advanced technology, but rather technology should be simplified and tailored to the needs of practitioners. Eligotech, a Dutch software development company providing big data solutions, offers a Hadoop-based data discovery tool, for people without any prior knowledge of this sophisticated software.

“Companies starting with big data spend at least four to five months on skills, resources and development before being able to generate value. Eligotech Harpoon helps companies instantly create value, saving time and cost. Analysts are spending 70% of their time on preparation, before they start analysing. Eligotech Harpoon can reduce this time dramatically.” (Eligotech corporate presentation)

As new technology emerges, it is initially mastered only by an elite few. However, as the tools become better and businesses adopt them, the technology gets diffused into the mainstream. Data analytics is increasingly penetrating business operations and government services; it cannot remain an area of “tech geeks” or holders of PhD degrees in mathematics. It should be further democratised, with new user-friendly software and applications set to conquer markets. On 16 September 2014, IBM, the global multinational technology and consulting corporation, announced a new cloud-based software called Watson Analytics. The idea is that a business person, not a data scientist, can type in questions to probe corporate data. Likewise, data-driven startups too work intensively to democratise data analytics.

“You can have data scientists who will do only certain manipulations, but future systems should be easy-to-operate, open for general audiences. Performing data analytics should be as simple as playing a video game. SynerScope is one of the companies that makes it happen. We are running experiments with 16-year-old kids, using our technology, and we obtain remarkably good results. The whole point of data analytics – it is about people, not computers.” (Jan-Kees Buenen, SynerScope)

“In a project three years ago, we tried to develop a new programming language, for people without programming skills. They would have been able to build visualisations and manipulate data, to sketch small solutions and innovate. So our intention was to lower the barrier and bring people with ideas closer to data.” (Fabien Girardin, Near Future Laboratory)

Despite these attempts to make big data software accessible for wider audiences, data scientists will continue to be in great demand. The increasing volumes of data that are becoming available, and the growing importance of data as a driver of competitive advantage will only intensify competition for top data talent across industries. As a policy implication, there is a clear need to develop new multidisciplinary educational programmes to train data scientists. These can be either MSc programmes or short-term postgraduate courses.
“A US-based company called Insight offers trainings in data analytics to people with a PhD degree in physics, mathematics, economics and related disciplines, who want to switch to data science. The company trains them for one month and helps find data analytics jobs. Smart people come out of the programme, and data companies eagerly hire them.”
(Eyal Amir, Parknav)

What is clear is that, at least in the near term, there appears to be a “crunch” in the market for data talent. Startups and smaller companies are likely to suffer most because they have to compete with large, established companies that can afford to pay higher salaries and offer more security for data talent. Startups are more likely to require “generalists” who are able to tackle a variety of analytical problems; yet, people with hybrid skillsets are particularly hard to find in the marketplace.

The regulatory environment: challenges and growth prospects
We also asked respondents about regulatory hurdles in their operations and their plans for the future, such as scaling up and international expansions. As a general background for their growth plans, many respondents expressed concerns and problems pertaining to the overall ecosystem and the role of government in their home countries in Europe, not necessarily inherent to the big data-driven business including access to capital, obtaining permits, rigid labour law and over administration of young enterprises.

When the range of problems facing startups is limited exclusively to those pertaining to the big data business, issues of Europe’s privacy and data protection come to the fore immediately. Most startups are law-abiding and act in good faith, but are puzzled by the existing rules and plead for regulatory clarity and simplicity.

“All rules surrounding data are very confusing. Can we store US-generated data in the EU, and vice versa? And what happens if data is treated in Europe, but by third country nationals in multinational companies?” (Asparuh Koev, Transmetrics)

“Privacy and data protection regulations are a big hurdle to innovative companies. They don’t have time and expertise to deal with that. Even if you have time, you won’t be able to understand legislation correctly.” (Paul Kremer, Mezuro)

And this issue achieves a multiplier effect when companies plan on using personal data.

“If you go into the big data problem, problems are scaled up fast. It is a trend lately that personal data needs to be protected. If you ask me, the point is completely missing – our data is not protected against those we should be afraid of and not available to those who need it and from whom we could benefit.” (Ivan Gligorijević, mBrainTrain)

Despite the sometimes fraught nature of the debate on data in Europe, a lot of data analytics is not “scary” or “unethical;” it is just an effort of companies to get to know their customers better, and to
provide them with more directly tailored services. The risk is that – in reaction to scandals like the 
Edward Snowden affair – Europeans will over-react and erect a regulatory structure that prevents 
data analytics businesses from flourishing in Europe, which, given the hugely important role that 
data-driven businesses are destined to play in the future, would be a mistake of historic dimensions.

To be sure, Europe needs a strong data protection regime – one that provides comfort and security 
to consumers that the personal data they share is secure and not subject to abuse. But it is equally 
clear that any data-sharing regime should be an enabling one, with rules that allow new businesses 
built on data analytics to emerge, a regime in which European startups can compete on an equal 
footing with startups elsewhere.24

Also, the generation gap seemingly plays its role in the data protection debate. The so-called 
“millennials,” the demographic cohort whose birth years range from the early 1980s to the 
early 2000s, are at the forefront of the digital revolution. Given their fluency and comfort with 
technology, millennials have a strongly positive view of how technology is affecting their lives.

“I do feel that the policymakers’ generation has overplayed the issue of data protection and 
privacy compared to the younger generation (millennials).” (Richard Edwards, DataShaka)

“I observe a generation problem. While an older generation is preoccupied with privacy, 
millennials are less concerned about it, e.g. in situations when apps use their personal data. 
In contrast, they are often surprised if companies don’t use their data to personalise services 
offered.” (Koen Havlik, Algoritmica)

Some aspects of the proposed general data protection regulation would help local businesses 
to thrive. For example, respondents complain that regulations on the access to personal data are 
heterogeneous across the EU, and this is especially burdensome for startups that do not have 
sufficient resources to study national legal frameworks of 28 EU member states.

“We find the personal data protection law 
tends to be rather uneven across the EU thus 
limiting scalability of some of our products.” 
(Fernando Alfaro, Madiva Solutions)

Another challenge is cultural, and can be 
related to adoption of new innovation solutions. 
Diffusion of innovations is a theory pioneered 
by Professor Everett Rogers in a book with the 
similar title first published in 1962. It posits that the speed of adoption is slow at the start, but 
more rapid as adoption increases. The same phenomenon is observed with products and services 
offered by data startups. Potential clients are hesitating to try and adopt them, and the client base 
remains limited at the starting phase.

“The speed of adoption may be a challenge. Organisations are very conservative when it comes 
to technology, not easy-going.” (Bruce Hellman, uMotif)

24 See, for example, Travis Korte, “Proposed EU Data Protection Regulations Could Impede Medical Research,” Center for Data Innovation, 21 
October 2014. Mr. Korte argues that the EU’s proposed general data protection regulation as currently written would “reduce researchers’ 
ability to conduct lifesaving research in cancer, infectious diseases and other areas.” He proposes that a “one-time consent” clause be 
added to the regulation, under which patients could effectively donate their anonymised medical data to science for research purposes.
“The novelty of our innovative product is the main obstacle. Our potential clients are not aware of its advantages. And hence, the decision-making speed is low. Our product transforms their routine business practices, and it is a difficult decision.” (Asparuh Koev, Transmetrics)

Some say the US offers a better regulatory environment and a better ecosystem. There are many examples of European (data-based) startups choosing the US as the place where they will grow and scale up. The good news though is that many are seeing the US as a temporary stage needed for future growth and a launch pad for their operations back into Europe.

“In 2014, we made a jump to the US, now from the US we are expanding back to Europe.” (Jan-Kees Buenen, SynerScope)

There are opposite cases too. BizEquity, a US-based provider of business valuations, is proudly rolling out its services in London.

“Europe has all ingredients for digital advancement. Nordic economies possess rich ecosystems with strong consumer orientation, and London leads in financial technology – fintech. Europe has a great opportunity to benefit from London’s leadership around financial technology.” (Michael Carter, BizEquity)

Parknav, a US-based street parking information provider, running algorithms using big data, is rapidly growing in Germany and planning a pan-European expansion.

“We expand to where we see our next customer. This is how we came to Europe. BMW, our most significant customer, attracted us to Bavaria. Germany is a very good place to start a new business [in the automotive-related sector]. The ecosystem has been very receptive to us. We plan to roll out our products in the top 10 cities in Germany by the end of the year; and to be in five more European countries by the end of the next year.” (Eyal Amir, Parknav)

Overall, scaling up is in the plans of many data-based startups – particularly as the same digital technologies many of them sell also facilitate the development of SMEs.

“On the product business, scaling up is no problem. We utilise cloud services that can scale up practically without a limit. We continuously experiment with new business opportunities that emerge in this rather new line of business.” (Markku Alanko, Ivorio)

Future prospects of data-driven business

Big data analytics offers vast business opportunities. Erik Brynjolfsson and his colleagues estimate that companies which inject big data analytics into their operations show productivity rates and profitability that are 5% to 6% higher than those of their peers.25 Another recent study found that more than 45% of companies have implemented a business intelligence or big data initiative in the 2010-2012 period.26 Another study, conducted in 2012, estimated that more than 90% of Fortune 500 companies would have at least one big data initiative underway within a year.27

---

27 Joe McKendrick, “Big Data Market Set to Explode This Year, But What is Big Data?” Smart Planet, 21 February 2012.
The universal consensus is that data analytics and big data will play a huge role in the economy of tomorrow. There is no way to stop the huge tide of data being amassed, and there is no real reason to try to do it. Businesses and governments need to embrace this transformational change now since Europe’s future growth hinges upon data-driven solutions and services.

“The use of big data right now is just a beginning. I think and I fear that companies that do not work with data to improve their products and services will not exist in 10 years.” (Peter Krekel, Clinical Graphics)

Even more data is becoming available as a result of the radically expanding Internet of Things, both globally and in Europe. The Internet of Things is a technology and market development based on the interconnection of everyday objects and applications among themselves. It will enable an ecosystem of smart application and services; and entrepreneurs feel new business opportunities are speedily opening up here.

“Originally, big data was about consumer data, now the shift is towards the Internet of Things; more data will be produced by machines. Many startups were active in consumer data, now they are moving to machine data.” (Koen Havlik, Algoritmica)

“The Internet of Things – it is going to do nothing more than generate a huge amount of data, and somebody will have to work with it and provide value and insights.” (Richard Edwards, DataShaka)

Thanks to their agility and creativity, startups are best positioned to benefit from the rise of big data and proliferation of data analytics. By their nature, they offer breeding grounds for new ideas, something that not all large companies can afford. Collaboration between startups and large established companies will remain a salient feature of the big data business. Such collaboration can be mutually beneficial – startups gain access to the resources of large companies, and in exchange established companies become exposed to novel ideas that would have never been possibly born inside the corporate walls.

“Agile, clever and innovative startups need to be taken into the mix to co-operate naturally with the large corporations and government organisations.” (Markku Alanko, Ivorio)

“Big businesses are waking up. They need to work with small and agile companies to unleash the potential of big data. Plenty of companies will fail, but others will stay. Lots of big companies are establishing accelerators of all kind. With startups, they can bring innovation from outside, as they cannot innovate from within for many systemic reasons.” (Richard Edwards, DataShaka)

“Marketing and advertising were the first functions within organisations to create value from big data technologies but the whole organisations should be data-driven... then we will have a real market and potential.” (Fabien Poulard, Dictanova)
National governments are responsible for creating environments conducive to data-driven entrepreneurship and facilitating evolution of ecosystems. Smart regulation is a key factor defining the success of data-driven entrepreneurship. The UK government is often cited as an excellent example in this regard.

“The UK takes a powerful stance in favour of entrepreneurs – it offers very supportive infrastructure, lower corporate tax rate, strong political support of 10 Downing Street to entrepreneurs, and business incubators, such as Innovation Warehouse.” (Michael Carter, BizEquity)

At the EU level, Europe’s data-enabled future will depend squarely on the completion of the digital single market, and much more dynamic decision-making inside of the EU. It suffices to say that the EU data protection directive was adopted in October 1995 at a time when Facebook or Twitter didn’t exist, and the information society directive in May 2001, a time when nobody even thought of YouTube.

“The Internet era fits poorly to country borders and legislation should be refactored significantly to match the market. The legislation emerges slowly and gets implemented locally even more slowly. The consensus is difficult to figure out on EU level and almost impossible on the global scale.” (Markku Alanko, Ivorio)

“The regulation evolves slower than the technology, and we must follow our own code of conduct to avoid potential liabilities. We have set up several measures to protect ourselves: ethical business code, technical security, appropriate partnerships with international law firms, strict service agreements, declarations to the privacy commissions in all countries we are active in, and concrete involvement in the European Commission’s workgroup on those matters.” (Laurent Kinet, Swan Insights)

While Europe faces challenges and is evidently lagging behind the US in the uptake of big data-driven technologies and solutions, the future is bright, the respondents feel.

“With big data, we are just getting started, and it will only be increasing. It is really exciting time for data-driven businesses!” (Richard Edwards, DataShaka)

“I am excited about business opportunities in Europe. Europe has a chance, specifically in fintech, to rival Silicon Valley and China.” (Michael Carter, BizEquity)

A fierce competition in the digital data-driven future will unfold not only for foreign direct investment from global multinational companies, but increasingly for talents and creative startups who will disrupt the status quo with their revolutionary data-powered products and services.

“Europe should welcome GSB – global small business – companies that are not restrained by national borders from their birth. Europe should embrace these companies, started in the US and elsewhere, and bring them to the EU where they establish their corporate HQ.” (Michael Carter, BizEquity)
Economic sectors
Big data and data analytics are not only allowing companies to improve and streamline their current operations, but also to offer many new services and create many new solutions in key parts of the economy.28

Financial services: Financial markets (banks, payment service providers, e-money companies) generate immense quantities of stock market and banking transaction data that can help companies increase trading opportunities or identify potentially fraudulent charges. The data that financial institutions accumulate enables easy and straightforward approval of loan requests based on simulated client risk profile, or may set flexible loan rates based on live market data. Profitability analysis can be easily executed too. Further, this data once anonymised and aggregated is a valuable input for new services and products, a source of innovation without crossing privacy concerns. Many new online services enter the “fintech” market; venture investment in global fintech tripled between 2008 and 2013 to almost $3 billion [or €2.3 billion], and is expected to reach $8 billion [or €6 billion] by 2018.29 Global players, such as Banco Bilbao Vizcaya Argentaria (BBVA) and HSBC, have developed in-house venture capital funds that invest directly in promising technology companies. The size of these funds can range anywhere from $50 million [or €37.6 million] to $250 million [or €188.2 million]. By investing early, banks gain early access to innovative solutions that can help reduce information technology costs and make their businesses more flexible to consumer need.30

Marketing, sales, retail: Big data is the biggest game-changing opportunity for marketing and sales, particularly because of the unprecedented array of insights into customer needs and behaviours it makes possible. Big data allows customer relationship management and customer acquisition, store location and layout, fraud detection and prevention, supply chain optimisation and demand signalling, dynamic pricing and target advertising. Retailers can analyse vast quantities of sales transaction data to uncover patterns in user behaviour, as well as monitor brand awareness and sentiment with social networking data.

Online services, web analytics: Browsing behaviour and text messages generated from streaming media, smartphones, and tablets represent treasures of knowledge about user behaviour and tastes. Internet companies initially exploited big data specifically to handle processing information at the Internet scale. Implementation of these analytical platforms is now viable for smaller online services companies to provide an edge over competitors for advertising, customer intelligence, capacity planning and more. E-commerce companies can benefit greatly from understanding customer behaviour and buying patterns via click stream, cohort analysis and other advanced analytics.

Logistics and manufacturing: Big data makes possible product research, engineering analytics, predictive maintenance, process and quality analysis, and distribution optimisation. Digital technologies enable substantial improvements of standard practices. Using radio frequency identification (RFID) sensors, handheld scanners, and on-board global position system (GPS) vehicle and shipment tracking, logistics and manufacturing operations produce vast quantities of information that can offer significant insight into route management, cost savings and operational efficiency. Data analytics make possible advanced manufacturing technology, or “smart

30 Idem.
manufacturing,” that is rapidly transforming the global competitive landscape. ICT-based solutions applied across the manufacturing process chain help make manufacturing efficient. Data analytics enables flexible manufacturing, optimal production rates and faster product customisation.

Research, new product development: Organisations have long recognised the importance of data analytics in new product development. They tap into the data generated from various sources for their product development processes, as such data offers huge potential to improve product performance, increase efficiency, contain costs and enhance customer experience. Many new product development organisations are critically analysing product failure data, service data, warranty data, historical design data, materials data and the like to extract information patterns that can be fed back into the new-product-development process.

Healthcare, life sciences: Healthcare is a sector where big data has all potential to bring about transformational change. Big data may find its application in pharmacogenomics, bioinformatics, pharmaceutical research and clinical outcomes research. Digital technologies enable real-time diagnostic monitoring analytics (improve service, predictive healthcare) in healthcare and analytics for on-demand clinical trial management (speed time to market) in pharmaceutical. Electronic medical records systems are some of the most data-intensive systems in the world and make sense of this data to provide patient treatment options. Big data is revolutionising medicine as we know it, and makes personalised medicine possible. Personalised medicine is a medical model that proposes the customisation of healthcare using molecular analysis - with medical decisions, practices, and/or products being tailored to the individual patient. Personalised medicine has the potential to help millions of people to enjoy longer and healthier lives. This is of crucial importance in Europe, which has one of the fastest ageing populations in the world.

Energy, utilities: Smart instrumentation, such as “smart grids” and power-line electronic sensors, sensors attached to machinery, oil pipelines, and equipment generate streams of incoming data that should be stored and analysed promptly to uncover and fix potential problems before they result in costly or even disastrous failures. Providers may use smart meter analytics to optimise rates, load-balance the grid and offer discounts, as well as operational modelling.

Transport: Big data provides new ways of gathering novel information about transport infrastructure from passenger and vehicle movements and allows for a shift from passive approaches to active crowdsourcing with innovative transport solutions. GPS systems enable drivers to inform others about congestion and incidents on the roads. This information is transferred to network operators in real-time, allowing for rapid responses. And intermodal car navigation proposes not only alternative routes based on real-time information using data from several different sources.

A 2011 report by McKinsey Global Institute, described the huge economic potential for data analytics in the public sector. If US healthcare, for example, were to use data more creatively and effectively to drive efficiency and quality, the sector could create more than $300 billion [or €234 billion at the 2012 exchange rate] in value every year. In the developed economies of Europe, government administrators could save more than €100 billion in operational efficiency improvements alone by using data more effectively, not including using data to reduce fraud and
errors and boost the collection of tax revenues. In total, it would lead to $250 billion [or €195 billion] of potential annual value to Europe’s public sector administration. Users of services enabled by personal location data could capture $600 billion [or €467 billion] in consumer surplus.31

Towards smart policy and smart regulation

As an interesting byproduct of the survey which underlies the conclusions in this study, we found that founders of data-driven startups are constantly busy with day-to-day operations, and may not have time or interest to be engaged or even monitor political decision-making process at the EU level in Brussels. Startups don’t have powerful lobby groups aiming to shape new legislation. They seek to comply with existing regulations but don’t necessarily understand their intricacies.

Data-driven startups need to be involved in democratic debate on new pieces of legislation at early stages. Once new regulations are enacted, it is too late, and it may impede their operations and disrupt their business models.

As our interviews show, data-driven startups operating across borders in the EU (or even in one country but with clients from other member states) are vocal supporters of a single European data protection regulation, fit for the digital age.

The European Commission plans to unify data protection with a single law, the general data protection regulation to replace the EU data protection directive adopted in 1995. Its draft was first unveiled in January 2012, and – as this policy brief goes to press – the European Parliament has adopted a position, but amendments are still being considered by the Council of Ministers. The expectation is to adopt the final language of the regulation in late 2014, with the rules coming into force by approximately 2016.

Informed consent constitutes a core element of the new general data protection regulation. Importantly, the regulation stipulates that organisations processing individuals’ personal data must obtain informed consent from individuals every time they want to use data for a purpose other than that for which it was originally collected. This principle would have far-reaching effects on various applications. For example, scientists might need to re-contact thousands of patients (some of whom may have subsequently deceased) for consent to use their data for medical research into diseases like cancer, infectious diseases and other areas.32 This leads some medical researchers to say that the regulation as written would deal a blow to European medical research in general, giving an advantage to other societies where the rules and regulations are less cumbersome.33 As Europe’s population is getting older, medical research and healthcare is one of the sectors that is set to grow in importance and gain dramatically from the application of data analytics, with ample opportunities for new data-driven startups.

32 Travis Korte, op. cit. An alternative proposal would allow “one-time consent,” in which the data provider would authorise use of his or her anonymised data for future, as-yet-unforeseen purposes, such as medical research, with a single, up-front waiver.
33 Ibid.
Another illustrative example pertains to cloud providers. A recent survey finds that only one in 100 cloud service providers in Europe are ready to meet the requirements of the proposed general data protection regulation.34

Against this backdrop, we propose 10 principles intended to serve as policy guidelines designed to improve the overall environment for data-driven startups.

1. Policy initiatives aimed at protecting European consumers against US corporations could have a counterproductive effect, damaging European startup entrepreneurship even more. This, in turn, could make Europe into a permanent “no-go” zone for data-driven businesses of all types, and further harm Europe’s economic growth prospects. Already, US-based startups in the data field have a distinct competitive advantage over their European rivals. In the US, datasets tend to be larger and more homogeneous, lacking the complexities generated in Europe by variations in standards and regulation.

2. Data privacy is a much more contentious issue in Europe than in the US. However, user behaviour, especially among young people eagerly embracing technology, is not fundamentally different in Europe and the US.35 Concerted efforts and confidence building are needed to give the industry (including startups) the enabling framework they need to continue developing important new services and social goods.

3. A smart regulatory environment and lean bureaucratic processes are a must. While the majority of businesses wish to act in good faith and comply with the regulations in place, the sheer complexity of the labyrinth of legal acts often makes regulatory compliance extremely hard if not impossible. Many are unaware of the intricacy of current regulations, not to mention the latest details of the on-going legislative processes.

4. Fragmentation of the European continent along national lines is extremely palpable when it comes to digital technologies. European companies need to overcome substantial hurdles in order to operate across digital national borders in the EU. This prevents many data-driven startups from gaining scale continent-wide. Policymakers should take tangible steps to allow companies to benefit from the scale that Europe could conceivably provide, perhaps with concrete policy initiatives such as the E-Corp, or “the 29th regime,” under which digital companies could operate under a unique legal regime covering all 28 EU countries. In this respect, the recent proposal of the European Commission on Single-Member Private Limited Liability Companies/Societas Unius Personae, or SUP, is a step in the right direction.36 A unified legal regime of this type would be transformational for European startups and also act as a powerful magnet for international entrepreneurs who might chose Europe as their destination of choice for global expansion.

35 European consumers are enthusiastic users of data-driven businesses, and particularly Facebook, Google search and Gmail. In the last quarter of 2013, Facebook had 193 million daily active European users and 282 million monthly active European users, more than it had in the US (the comparative figures for North America, which includes the US and Canada, are 147 million active daily users and 281 million monthly users). The Google search engine has an overwhelming 90% to 96% market share in most European countries, more than in the US. For more, see Facebook Annual Report.
Entrepreneurs urgently call for the harmonisation of regulations on the use of personal data (to speed up the adoption of the general data protection regulation), and for making those rules fit for the digital age. European businesses need to operate with a single, transparent and straightforward data protection regime, rather than a patchwork of 28 national regimes.

Open data is a vital source of data for startups, and thus the open data movement should be promoted with the massive opening up of public data resources for use and re-use as a tangible first step. The EU Open Data Portal is an excellent example of an action at European level; likewise national open data portals in France, the Netherlands and the UK set high standards. These countries are the top three European leaders in the UN e-government survey 2014. The French government’s 16 September 2014 decision to create a post of chief data officer who will work directly with the secretary of state and prime minister is a forceful message that open data is a strategic priority in France. Europe needs to launch a comprehensive study on how startups can use open government data to generate new business and develop new products and services, similar to The Open Data 500 in the US.

Policymakers should facilitate the formation and evolution of a coherent data ecosystem bringing together large software firms, SMEs and startups, the research and academic communities, and other organisations in data-intensive sectors. Collaboration between data-driven startups and established large corporate players should be promoted. Large corporations possessing massive datasets (in the telecom, finance and other broad sectors, for example) could be encouraged to release them in an anonymised and aggregated form without revealing any corporate secrets and subject to all privacy requirements. It will provide new opportunities for growth and innovation, including for data-driven startups.

Data portability – the ability for people to re-use data across interoperable applications – should be facilitated. While big data itself is inherently resistant to standardisation, common standards and algorithms for data analytics should be promoted, as they will contribute to the portability of data analytics results.

Facilitate access to financing for startups and scale-ups through a regulatory framework that creates a secure and level playing field for crowdfunding, venture capital and other forms of new enterprise financing. European Commission President Jean-Claude Juncker’s proposal for a Capital Markets Union would be a bold step in the right direction. The plan, which is included in President Juncker’s Political Guidelines for the Next European Commission, would ease access to finance for companies and infrastructure projects and “reduce our very high dependence on bank funding,” as President Juncker explained to the European Parliament at his confirmation hearings. Moves like these to improve and deepen access to capital for European companies are urgently needed, particularly for technology-intensive startups.

37 At the EU level, several initiatives have been implemented, including the Directive on Re-Use of Public Sector Information. In December 2011, EU legislators adopted the Open Data Package, including the amended version of Directive 2003/98/EC.
Closing the job skills gap and addressing skills shortages is of paramount importance. Data management skills are among the most sought-after competences. Startups need data managers and business intelligence analysts to implement data-driven business models. “Data scientists,” combining analytical and statistical skills with an appropriate level of business understanding, will be in high demand. Thomas H. Davenport and D.J. Patil refer to this profession as “the sexiest job of the 21st century.”41 European government can design short-term data analytics training programmes, or support companies offering services of converting talented people to data scientists in a relatively short amount of time.

Europe has all the potential to seize the vast opportunities generated by the data-driven revolution, and our research shows that many startups are seizing it. Sustained, concerted efforts of European and national policymakers, industry leaders and the entrepreneurial community are of paramount importance to drive this movement forward, helping Europe to lead in area where tomorrow’s global economic leadership will surely lie. In the words of President Juncker, “The Commission needs to play its role in ensuring that promising new developments such as the cloud, the Internet of Things and big data can thrive in Europe and that citizens, innovative web entrepreneurs and other businesses can take full advantage of their potential.”42

41 Davenport and Patil, op cit.
The author would particularly like to thank the following individuals for incisive commentary, outstanding research and excellent collaboration during the preparation of this policy brief, which was very much a team project, compiled in the spirit of open collaboration, “team sourcing” and knowledge sharing:

- Marco Bressan, chairman and CEO, BBVA data and analytics
- Carmen Cuesta Sainz, senior analyst, BBVA research
- Tatiana Issaeva, venture capital investment manager, European Investment Fund
- Juan Mateos-Garcia, research fellow in economics and creative economy, Nesta
- Telmo Pérez Luaces, vice-president business development and strategy, Orange
- Thanh Nguyen, deputy head of Brussels office, Orange
- Eduardo Salido Cornejo, public affairs and policy manager, Telefónica
- Jonny Shipp, head of digital confidence, Telefónica

Special thanks as well to Jacky Abitbol (Orange), Francisco Javier Arias Marin (BBVA), Nathalie Boulanger (Orange), Christopher Haley (Nesta), Paul Hofheinz (the Lisbon Council), Frédéric Michel (Telefónica), Richard Poston (Telefónica), Marjut Santoni (EIF), Gustavo Vinacua (BBVA), Stian Westlake (Nesta) and the project teams at the Lisbon Council and Nesta.
Appendix: Data-Driven Startups

The companies that participated in our research form a diverse and rich sample.

- **Algoritmica**, a company based in the Netherlands, combines machine-learning algorithms with the power of supercomputers to build unparalleled predictive models for marketing, risk, fraud, supply chains and maintenance. Its website is [www.algoritmica.nl](http://www.algoritmica.nl).

- **Big Data Scoring** is a European provider of credit scoring solutions based on social media. The company was founded in 2013 in Tallinn and has offices in Estonia, Finland and Poland. Its services are aimed at banks and consumer lending companies. Its website is [www.bigdatascoring.com](http://www.bigdatascoring.com).

- **BizEquity**, a US-based startup rapidly expanding in the UK, has developed an online business valuation engine to help business owners and entrepreneurs receive an estimate of what their business could potentially be worth. Its website is [www.bizequity.com](http://www.bizequity.com).

- **CitizenMe** is a UK-based personal identity management service helping people take back control of their digital profiles on social networks (Facebook, LinkedIn, Twitter), apps and all other places across the web. It helps people make the “virtual you” appear in the way they want it to. Its website is [www.citizenme.com](http://www.citizenme.com).

- **ClinicalGraphics**, a spin-off from Delft University of Technology in the Netherlands, creates visual interactive motion simulations for hip impingement and dysplasia in a 3D PDF format to support orthopaedic treatment decisions. Its website is [www.clinicalgraphics.com](http://www.clinicalgraphics.com).

- **DataShaka**, a UK-based company, offers a unification platform, enabling customers to combine data, no matter the source or format, all powered by its TCSV algorithm, standing for Time, Context, Signal and Value. Its website is [www.datashaka.com](http://www.datashaka.com).

- **Dictanova**, a France-based company, is a startup specialising in text mining and qualitative analysis for marketing departments. It helps marketers monitor and develop the perception of their brand on the Internet and optimise their relationships with their customers. Its website is [www.dictanova.com](http://www.dictanova.com).

- **DTP Solutionpath**, a UK-based company, offers a technology that creates contextual relationships, can automatically catalogue, add metadata in context of organisations practices (folksonomy) and retrieve conceptually and contextually relevant data. Its website is [www.solutionpath.co.uk](http://www.solutionpath.co.uk).

- **Eligotech** is a software development company in the Netherlands, focused on creating products that solve big data issues. Eligotech’s mission is to make the introduction and usage of the most advanced big data technologies easy, secure and affordable for any company. Its website is [www.eligotech.com](http://www.eligotech.com).

- **GoDataDriven** is a young big data startup based in the Netherlands. Its objective is to help organisations create end-to-end data driven solutions. The company is a big data service provider offering a wide range of solutions that can be tailored to the needs of the company. Its website is [www.godatadriven.com](http://www.godatadriven.com).
Ivorio is a Finland-based consultancy company specialising in the concepts of big data and the key implementation technologies. Its goal is to clarify the potential of big data in its customers’ business with tailored training, mentoring and in-depth analysis. Its website is www.ivorio.fi.

Madiva Solutions, a Spain-based company, aggregates information and develop proprietary algorithms to create tailor-made solutions in customer service, customer relationship management, segmentation and knowledge management. Its website is www.madiva.com.

mBrainTrain is a Serbia-based company aiming to improve medical services and quality of life through electroencephalography software and hardware. It develops a fully mobile, wearable device for recording and analysing electrical brain activity. Its website is www.mbraintrain.com.

Mezuro, a company based in the Netherlands, develops reporting systems for mobility measurements; it counts the number of people and vehicle movements in an area on the basis of anonymous statistic data from mobile communications networks and counting cameras. Its website is www.mezuro.com.

Near Future Laboratory, a company launched in Spain and having offices in Barcelona, Geneva, San Francisco and Venice Beach, aims to understand how imaginations and hypothesis become materialised to swerve the present into new, more habitable near future worlds. Its practice involves working closely with creative, thoughtful experts within various domains of work. Its website is www.nearfuturelaboratory.com.

Parknav is an information technology company specialising in location-based intelligence regarding available parking space. Parknav gives complete coverage of Chicago and San Francisco in the US, and Munich and Hannover in Germany, including all free spots, all zoned (permit) spots and also all pay (metered spots). The new concept calculates with new principles based on big data instead of the known park-sharing methods. Its website is www.parknav.com.

Swan Insights is a leading Belgian data-driven startup whose approach to big data fuses corporate and external data to generate hidden insights underlying people’ and groups’ social behaviours. Its website is www.swaninsights.com.

SynerScope founded as a high-tech spin-off company of Eindhoven University of Technology in The Netherlands, provides advanced business intelligence/big data analysis software that directly allows actual domain experts and analysts to make sense of their big data. Its website is www.synerscope.com.

Transmetrics is an innovative Bulgaria-based startup that is solving the biggest problem of the cargo transport industry – capacity utilisation. It is tackling the problem by using the power of big data and predictive analytics. Its website is www.transmetrics.eu.

uMotif is a UK-based startup offering a simple and engaging software platform that puts people at the heart of their care. Using its web and mobile apps, people track and monitor their health and choose to share their data with their clinicians to improve health outcomes. Its website is www.umotif.com.
About the European Digital Forum

The European Digital Forum is a think tank dedicated to empowering web entrepreneurs and growing Europe’s digital economy. The initiative is led by the Lisbon Council, a European think tank based in Brussels, and Nesta, the United Kingdom’s innovation foundation, in collaboration with the European Commission’s Startup Europe initiative. The European Digital Forum was launched at the World Economic Forum in January 2014 as a vehicle to intellectually accompany the 22-point action plan put forth in the Startup Manifesto (www.startupmanifesto.eu) written by the Leaders Club, an independent group of founders of world-leading technology companies based in Europe, including Atomico, HackFwd, Rovio, Seedcamp, Spotify, Tech City Investment Organisation (TCIO), Tuenti and The Next Web. In the manifesto, which was drafted to spur discussion on improving Europe’s startup ecosystem and digital-era performance, the European tech leaders proposed establishing a permanent independent think tank to explore and elaborate a more decisive approach to startups, an invitation which was seized and carried forward by the Lisbon Council and Nesta in 2014. Among the founding partners of the initiative are the European Investment Fund (EIF), Banco Bilbao Vizcaya Argentaria (BBVA), Orange and Telefónica. Follow the European Digital Forum on twitter at www.twitter.com/edf_eu.

www.europeandigitalforum.eu