Deutsche Bank
Research

Current Issues
Digital economy and structural change

Big data
The untamed force

May 5, 2014

Authors
Thomas F. Dapp
+49 69 910-31752
thomas-frank.dapp@db.com

Veronika Heine

Editor
Lars Slomka

Deutsche Bank AG
Deutsche Bank Research
Frankfurt am Main
Germany
E-mail: marketing.dbr@db.com
Fax: +49 69 910-31877

www.dbresearch.com

DB Research Management
Ralf Hoffmann

Big data is increasingly becoming a factor in production, market competitiveness and, therefore, growth. Cutting-edge analysis technologies are making inroads into all areas of life and changing our day-to-day existence. Sensor technology, biometric identification and the general trend towards a convergence of information and communication technologies are driving the big data movement.

Huge challenges must be overcome if the benefits are to be leveraged effectively. Matters of concern alongside increasing volumes of data, varying data structures and real-time processing include data security, data privacy policies that are in urgent need of reform and the rising quality expectations of the stakeholders. There is a widespread lack of suitable strategies to respond to the digital revolution.

Data has an economic value. The appetite of many stakeholders for data, including personal data, is growing. Internet users themselves could become the traded commodity, without them ever being aware of this or experiencing any financial benefit.

The opportunities for business, academia and politics are manifold: they range from increases in productivity, reduced costs and optimised customer communications to better disease forecasting and scientific modelling. In addition, more and more public authorities are offering anonymised data sets as a means of generating new products, services and business models.

The risks are not to be underestimated: many people complain not only of a growing feeling of information overload but also of an increasing loss of data sovereignty. Any number of stakeholders are compiling, storing and analysing data, including personal data, for any number of different reasons. If only some of them misuse this data, trust in digital channels could be broken, or at least significantly reduced. This would deal a fatal blow to innovation and growth, and not only in the still relatively young field of research into information technology.

The big data movement cannot be stopped. It is now a question of putting in place the necessary regulatory framework to allow these state-of-the-art methods and the technology that underpins them to properly flourish. With their success in the fields of sensor technology and biometrics as well as data protection, Germany and the European Union now have the opportunity to play a leading international role in the development of secure IT infrastructures.

Big data – the untamed force

1. Big data – A next logical stage of evolution for the internet

"Technology is neither good nor bad; nor is it neutral.[…] Technology's interaction with the social ecology is such that technological developments frequently have environmental, social and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves." [Melvin Kranzberg, 1986]
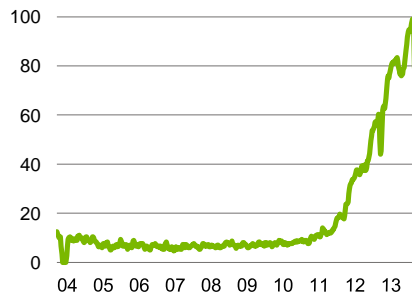
Could there possibly be a correlation between people buying books and shoes from the online retailer Amazon at a certain time in the evening? Are credit card companies able to predict which of their customers may be heading for a marriage crisis, solely on the basis of individual transactions? Might it be possible to use GPS tracking of mobile devices in combination with near-real-time analysis of microblogging posts to predict and prevent or at least mitigate an impending epidemic? In the future, could an internet-enabled e-toothbrush send information about your oral hygiene straight to your dentist so that they can see if you need to make an appointment?

These examples may seem somewhat outlandish at first, but the controversial and increasingly widespread discussion about the mega topic of 'big data' puts them in a completely new light. Using sensors, a multitude of data sets and specific algorithms, automatic predictions could soon be made about particular behavioural tendencies (and not just online) on the basis of simple correlations. Hidden in the ocean of data could be any number of unforeseen, surprising and potentially valuable correlations that we can only begin to imagine today.
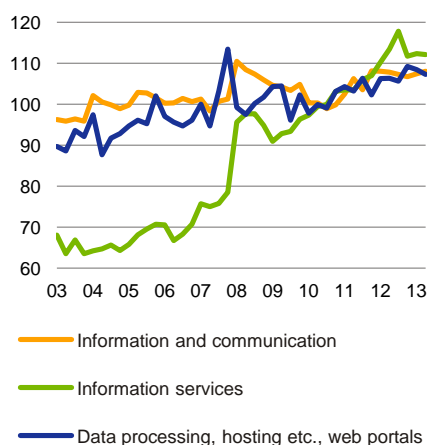
The way in which people think about data and data analysis will gradually change as well, in addition to the technological possibilities. Thanks to the latest internet technologies, the potential for harnessing all that can be measured and analysed using solid data, intelligent sensors and filtering has never been as promising and lucrative as today, at the dawn of the digital era.

We are in the midst of the digital revolution. The importance of web-based technologies is growing in all sectors of the economy. Almost all commercial transactions, and many private ones too, are now carried out digitally, which brings down the cost. Digitisation is changing society and people's day-to-day economic lives, and stakeholders from business, academia and politics are responding to this – albeit at different speeds. The ever-growing economic benefit behind this digital development is derived from individual data points, which are increasingly being queried in real time.[1]

Big data is the new, hotly debated topic that follows a series of logical stages in the development of the internet, such as individualisation, the relocation of data to the cloud and the rapidly growing demand for digital mobility. It bridges the gap to what has evolved before. In principle, the idea is to combine different volumes of data with new data sets and to identify any patterns in this aggregated data using intelligent software, with the ultimate aim of drawing the right (and wherever possible, lucrative) conclusions from the findings. Once they have been compiled, primary data sets can be analysed any number of times for different purposes and for different stakeholders. The data functions as a driver of innovation, creativity and out-of-the-box thinking, and in an ideal world results in new business ideas, products or services.

However, it can and will lead to widespread concerns and fears, because data sovereignty, i.e. the degree of control that people have over their own information, can quickly be compromised, as many examples show.

At the end of the big data process, it is important to draw the right conclusions from the combination of technology and methodology. This poses another challenge for the decision-makers, as this is where two worlds collide: the bits,

A new way of thinking about data

Web searches for the term 'big data'   1

Worldwide interest, 2004-present, smoothed

Source: Google Trends (www.google.de/trends/)

ICT revenue growth   2

Germany, 2010=100, adjusted for working days and seasons

— Information and communication
— Information services
— Data processing, hosting etc., web portals

Source: German Federal Statistical Office

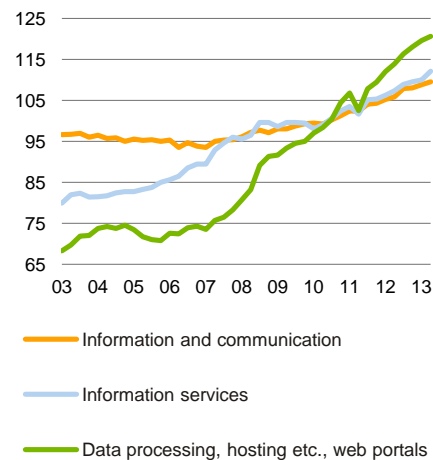[1] Identifying patterns or useful information from large batches of data is known as data mining.

Big data – the untamed force

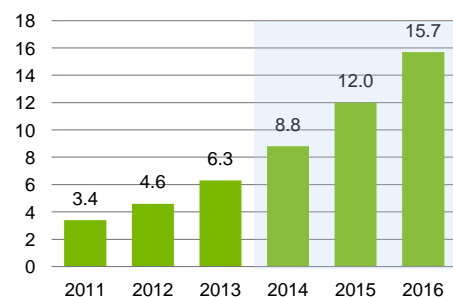**Employment in the service sector (ICT)** 3

Germany, 2010=100

- ——— Information and communication
- ——— Information services
- ——— Data processing, hosting etc., web portals

Source: German Federal Statistical Office

The desire of man to quantify the world

Global market potential of big data 4

EUR bn

Sources: Bitkom, Experton Group

bytes and algorithms on the one side and us, the people, on the other, with our instincts, experience and intuition. What counts more when it comes to making a decision? The results drawn from the statistics or a person's long-standing experience or gut instinct? The potential that computer-based algorithms have for extracting useful information from the combined data sets is undoubtedly still in its infancy and there is plenty of room for improvement. The hope remains, however, that despite the technological progress and the barely imaginable pace of development in this field, a harmonious relationship between man and machine will emerge. Leaving aside the fact that data privacy regulations in this still young international field of research have so far proved inadequate, surely only very few people would like to live in a world in which decision-making processes are left solely to programmed algorithms, even to those that have the capacity to learn. On the other hand, of course, information and communication technologies are able to increase efficiency, reduce costs and minimise human error – quite a trade-off.

This far-from-trivial challenge is a growing concern for stakeholders from business, academia, politics and society. The commercial importance of real-time data, such as financial and trading transaction, sensor/measurement, health and social network data, is growing all the time, as is the impact this has on business processes in a globally connected economy and on people's personal lives and social environment. More and more data is being generated by different stakeholders that can (or could be) filtered, analysed and used for different purposes several times over – and this a trend that is growing exponentially.

Although we are only at the start of the big data revolution, the technology and techniques at our disposal mean we are heading towards a situation in the long term in which people's actions, activities and moods can be made more and more measurable. This means turning data into a form that can be read by a machine so that it can be combined, perhaps with other data sets, and analysed. Thanks to the possibilities presented by sensor technology and biometric identification, we are getting closer to being able to quantify the world and the individuals that inhabit it – an aspiration that has existed for as long as human thought. It also brings us closer to the visions of humanoid robotics and adaptive artificial intelligence, as has been shown in countless science fiction films and now could become reality for the mass market too.

As with every new and (r)evolutionary movement, there are opportunities as well as risks. Both should be subject to a broad and transparent discussion at this early stage of development. From a German and European perspective, it is important for reasons of competitiveness to create a favourable environment now that will stimulate the potential of the big data movement. Meanwhile, the scope for legally and morally questionable practices involving the misuse of data must be limited by legislation on data protection that applies throughout Europe and ideally further afield.

Which steps are taking us closer to big data? The following chapter introduces the different stakeholders and explains the basic characteristics of the data. There will also be a focus on the sources and drivers of the growing flood of data as well as a number of examples that illustrate these. Chapter three puts the spotlight on sensors and biometrics. Both technologies are becoming increasingly suitable for the mass market and hold huge and lucrative potential for growth. The convergence of information and communication technologies, as a cross-sectoral trend, is generating additional impetus and scope for experimentation and for new business models, innovative products and services. The role of the digital ecosystem in the big data discussion is explored in chapter four. Certain large internet companies are expanding their core competencies with the aid of modern technologies and techniques in the field of data analysis. This increases not only the spectrum of their products and
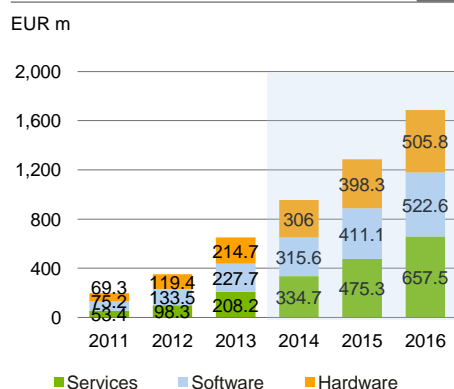
services but also the degree to which their customers are 'locked-in' and the costs of switching to another provider. Data has a commercial value. All kinds of data sets can be used any number of times by different stakeholders for different purposes; they can be monetised, but they can also be abused or misused. It is well within the realms of possibility to create digital profiles of people that are uncannily (and even frighteningly) realistic – and for the most part without obtaining their consent. The excessive data-gathering activities of individual stakeholders, a subject that can be hard for the layman to understand, are discussed in chapter five.

Chapter six explores the data protection policies that are inadequate for the big data movement but are fundamental to its future, and proposes several solutions. Chapter seven gives a brief insight into various big data projects in the realms of business, academia and public administration. The limits of the big data movement, most notably the increasing complexity caused by merging different data sets, are the subject of chapter eight. Chapter nine closes this study with a brief summary and a look ahead to where the journey might take us in the coming years.

2. Stakeholders, data characteristics, sources and drivers

"I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?" [Hal Varian, Google's Chief Economist on the subject of statistics and data, January 2009]

Big data is more than just IT: Stakeholders in all kinds of sectors

Many decision-makers have recognised that big data is no longer purely the preserve of IT. Big data is instead becoming a movement that brings together cutting-edge internet technologies and analysis techniques in order for large, extendable and above all differently structured data sets to be captured, stored and analysed. This gives big data a broad, international dimension with different knowledge-based outcomes and expectations with regard to increasing growth and efficiency. But above all, big data provides scope for experimentation, innovation and creativity, offers a wealth of potential new data combinations and is therefore ideal for discovering unexpected correlations. It could be used to create new business models, products and services and to drive innovation.

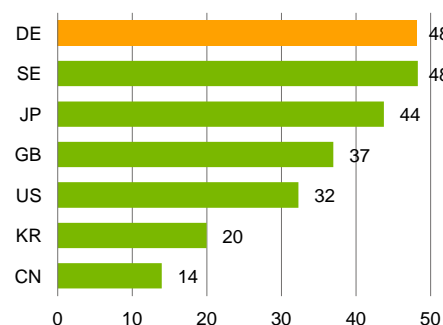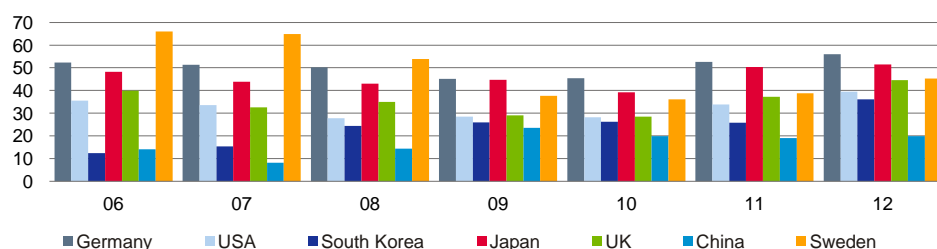Market potential of big data in Germany | 5

EUR m

Sources: Bitkom, Experton Group

DE and SE with the highest success rates for ICT patent applications | 6

%, average percentage success rate of patent applications

Source: EPO

ICT* patents | 7

%, average percentage success rate of patent applications

* EPO definition: audio-visual technology, telecommunications, digital communication, basic communications processes, computer

Source: EPO

Because of Germany's scarcity of natural resources and the strength of its research-intensive products and services, the big data movement clearly has the potential to become a stable comparative advantage for the German
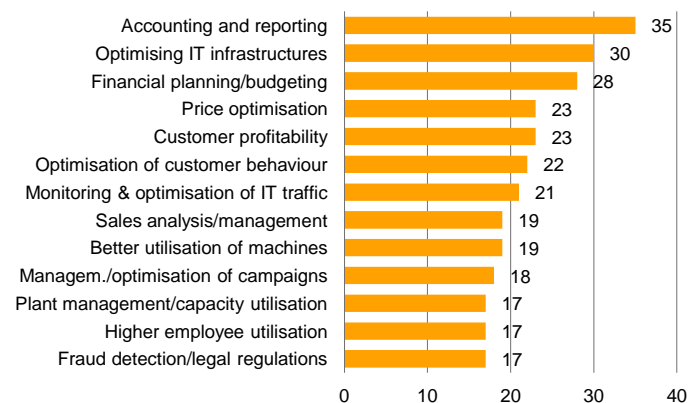
economy. Big data can become a factor in production and competitiveness that will open up new possibilities for value creation and ease the path for (r)evolutionary movements similar to that of the mass use of the internet. Above all, providers of new, high-performance data analysis tools are expecting long-term increases in efficiency and, in particular, lucrative revenue opportunities, for example through the collection and analysis of detailed customer-specific information: in short, big data ultimately means big business.

In addition to the IT sector, stakeholders from the worlds of business, academia and public administration are increasingly being confronted with big data. Mining the mountains of data that are produced by various different sources has obvious economic benefits but is also being used more and more to address social and humanitarian (e.g. medical) problems, as an example in chapter seven illustrates. However, the success of these efforts hinges entirely on the modern technologies and developments being underpinned by a mandatory legal framework. This needs to meet the requirements of the data protection policies on a European (and, ideally, international) level and to put information privacy at the forefront of discussion.

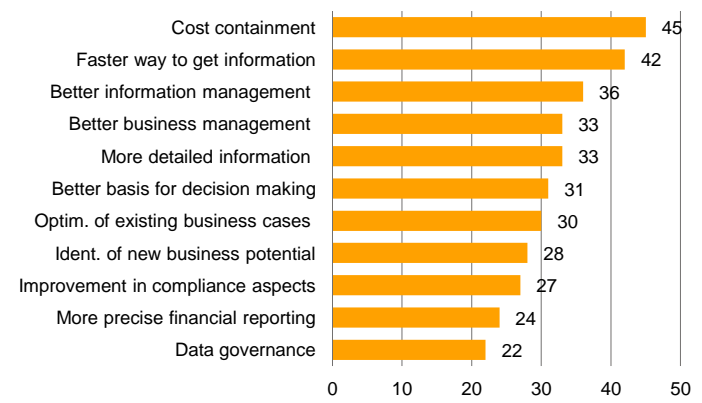Areas of business that stand to benefit the most 8

% of respondents* (n=254), 2012

Accounting and reporting	35
Optimising IT infrastructures	30
Financial planning/budgeting	28
Price optimisation	23
Customer profitability	23
Optimisation of customer behaviour	22
Monitoring & optimisation of IT traffic	21
Sales analysis/management	19
Better utilisation of machines	19
Managem./optimisation of campaigns	18
Plant management/capacity utilisation	17
Higher employee utilisation	17
Fraud detection/legal regulations	17

*multiple answers permitted

Source: IDC

Corporate potential of big data 9

% of respondents* (n=254), 2012

Cost containment	45
Faster way to get information	42
Better information management	36
Better business management	33
More detailed information	33
Better basis for decision making	31
Optim. of existing business cases	30
Ident. of new business potential	28
Improvement in compliance aspects	27
More precise financial reporting	24
Data governance	22

*multiple answers permitted

Source: IDC

The characteristics of data: Complex, heterogeneous but of huge value

The defining properties of big data, known as the three Vs, are introduced in the following section using a number of examples and illustrations:

— Volume

The first criterion refers to the volumes of data that occur in research and business and in the private and public spheres. The relentless progress of digitisation means that virtually all aspects of modern life are involved. In the private area alone, people store and archive an enormous volume of data that runs into the three-digit gigabyte range. Included in this are digital photos, documents, spreadsheets, e-mails, music files and film downloads. Companies, research institutes and public authorities also store, archive and analyse masses of data, but this is done on a far greater scale.

Big data – the untamed force

More bits and bytes than grains of sand on the beach

The storage capacity of mobile devices (smartphones, tablets, e-readers) or laptops is usually measured in gigabytes, a quantity of data we can still just about get our head around.[2] However, businesses, research institutes and even intelligence agencies that generate, manage and analyse vast sets of data on a daily basis are now using terms such as terabyte, petabyte, exabyte and zettabyte. Last year, 1.8 zettabytes of data were produced worldwide. One zettabyte is equal to one sextillion bytes and would be followed by 21 zeroes (1,000,000,000,000,000,000,000).

Units of measurement for data at a glance				10
1 Byte	=	8 Bits	=	1
1 Kilobyte (kB)	=	10^3 Byte	=	1,000
1 Megabyte (MG)	=	10^6 Byte	=	1,000,000
1 Gigabyte (GB)	=	10^9 Byte	=	1,000,000,000
1 Terabyte (TB)	=	10^{12} Byte	=	1,000,000,000,000
1 Petabyte (PB)	=	10^{15} Byte	=	1,000,000,000,000,000
1 Exabyte (EB)	=	10^{18} Byte	=	1,000,000,000,000,000,000
1 Zettabyte (ZB)	=	10^{21} Byte	=	1,000,000,000,000,000,000,000
1 Yottabyte (YB)	=	10^{24} Byte	=	1,000,000,000,000,000,000,000,000

Source: author's illustration

Data traffic on mobile, web-enabled devices 11

Petabytes per month
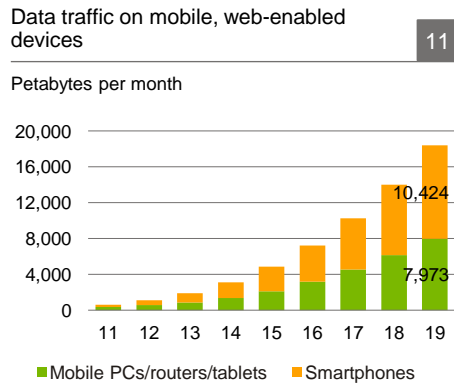
Source: Ericsson 2013

"Experts have estimated that mankind generated around five billion gigabytes of data from the beginning of recorded history to the year 2003. […] In 2011 the same volume of data – 4.4 exabytes – was amassed every 48 hours."[3] The trend has continued: with more powerful computers, cheaper storage media and smarter algorithms, these volumes of data occurred every ten minutes in 2013. So if you believe the experts, the volume is doubling every two years. The 'digital haystack' is getting bigger and bigger, making it more and more difficult to look for the needles. This justifiably raises the question of what the different stakeholders are actually looking for in this ever-growing mass of data. It is particularly interesting to discover who exactly has access to which data sets, what analysis techniques they are using and for what purpose are they doing this. The challenge often lies in the incredible rate at which data is produced, as companies would ideally like to analyse this in real time. This brings us on to the second characteristic:

— Velocity

Whereas data in the analogue era was generated, captured and analysed at manageable frequencies, the data flows in today's digital era are being produced around the clock and all over the world, in part because of the multitude of interconnected sensors and the merging of information and communication technologies (ICT). In order to meet the requirements of time-critical business and decision-making processes, it is becoming increasingly important to integrate (or analyse) data flows from both internal and external sources in real time. These are exactly the areas in which many decision-makers are hoping to leverage huge gains in competitiveness and efficiency. Science, too, is benefiting from real-time data, for example, when identifying pathogens in time for preventive steps to be taken before a disease becomes too widespread.

---

[2] Two gigabytes of storage space can accommodate around 360 high-quality MP3s, one 90-minute HD film and approximately 150 photos at five-megapixel resolution.

[3] Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. (Small data, big impact. Big data in a nutshell.) Landesanstalt für Medien Nordrhein-Westfalen (LfM). Düsseldorf.

— Variety

The final V concerns the variety of different data types and structures, which as far as possible must be processed in a standardised manner and 'harmonised' so that they can be properly combined. A basic distinction can be made between structured, semi-structured and unstructured data.

Whereas previously we mostly processed structured information in relational databases, newer data sets, e.g. from social networks, are increasingly structure-free, making conventional databases far less suitable for extracting value. Master customer data is an example of a structured format and includes gender, date of birth and address. Images, videos and audio files are types of unstructured data. E-mails, on the other hand, are categorised as semi-structured because the header information contains structured data, i.e. sender, recipient and subject. The bulk of the e-mail is unstructured, however, as the body of the message has no predefined structure. Experts estimate that nowadays only 15 % of data is structured, with around 85 % regarded as unstructured.[4]

The current challenge for IT network architects lies in making heterogeneous data formats from different data sources compatible with each other so that they can be stitched together for analysis. Integrating this data, however, is complex. Information ultimately has to be converted into machine-readable data and it can originate from a bewildering array of sources, including images, videos, MP3 files, streaming services, microblogging media, blogs, forums, search engines, click protocols, e-mails, internet telephony, digital correspondence and/or sensors. Converting audio, video and image files into standardised machine-readable data is a particular challenge. The underlying algorithms have to be able to understand and transcribe the language(s), recognise faces and corporate logos and even identify digital content that is protected by copyright.[5]

In the discussion initiated by Tim Berners-Lee[6] about using the semantic web[7] (Web 3.0) to develop the Internet of Things (ubiquitous computing), the call is put out for data to be tagged with semantic information so that it can be read by other machines and be properly classified. Examples: does a mention of the word 'Golf' refer to the sport or the model of car, or could it even be a misspelling of 'gulf' as in Gulf of Mexico? How would the e-toothbrush use data from its sensors to convert evidence of plaque and tartar build-ups into a machine-readable language for analysis at your dental practice?

Subjective comments made in written, spoken or video content are particularly interesting in this context. PR companies, election campaign agencies and marketing departments all have an interest in recording people's sentiments and opinions on particular subjects in near real time. In order for these often emotional responses to products, brands or events to be made machine-readable, intelligent analysis tools and algorithms are needed to identify the pertinent messages. Because, technologically speaking, this is a huge challenge and because further research is required, this will undoubtedly become the supreme discipline in big data.

Big data and the US election campaign 12

Several state-of-the-art big data analysis tools were used in Barack Obama's re-election campaign. They provided answers to the questions of how to widen to the impact of the campaign and how to efficiently distinguish people who might potentially vote Democratic from those who were already 'lost'. Obama's team used socio-demographic data such as consumer behaviour (information from loyalty programmes, shopping behaviour, online click patterns, etc.) and data from social networks (e.g. Facebook profiles). All this data was correlated with the electoral rolls in order to identify any patterns in people who were not intending to go the polls but if they did would vote for Obama. With the aid of various forecasting instruments (predictive analytics) up to 100 variables were fed into the system in order to predict voting behaviour. Data privacy experts were critical, arguing that many of these practices fell into a regulatory grey area and that individuals' information privacy rights were being infringed. However, because the data privacy laws in the USA are not as strict as in Germany or Europe, Obama's campaigners were able to combine and analyse a wide range of data sets relating to US voters.

Source: http://www.predictiveanalyticsworld.com/patimes/eric-siegel-explains-how-the-obama-2012-campaign-used-predictive-analytics-to-influence-voters-2

Worldwide smartphone subscriptions 13

Bn

	10	11	12	13	14	15	16	17	18	19
	0.5	0.9	1.3	1.9	2.6	3.3	3.9	4.5	5.1	5.6

Source: Ericsson Mobility Report 2013

---

[4] TNS Infratest GmbH – Technology business segment: Quo Vadis Big Data – Herausforderungen – Erfahrungen – Lösungsansätze. (What next for big data – challenges – experiences – approaches.) 2012.

[5] Cf.: Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. (Small data, big impact. Big data in a nutshell.) Page 12.

[6] Inventor of the world wide web.

[7] Whereas the internet offers a means for data to be linked, the semantic web provides a way of linking information at the level of meaning. Data in a semantic web is structured in a format that enables computers to process information based on the meaning of its content. A semantic web would also allow computers to extract knowledge from the mass of global information and, more importantly, to generate new knowledge.
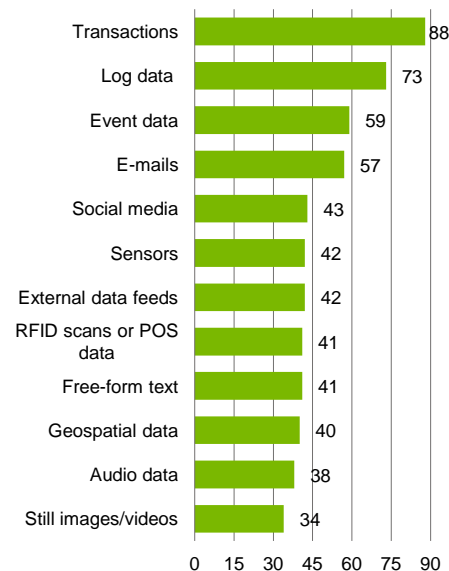
Big data – the untamed force
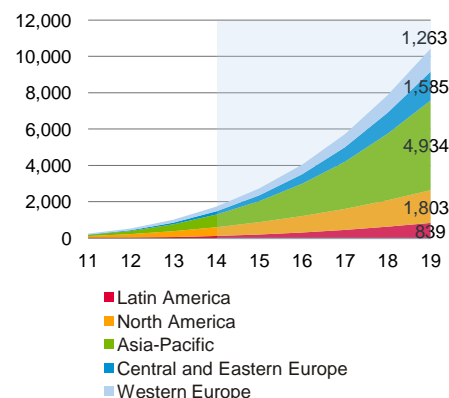
Experts are now talking about a further V in addition to the usual three characteristics:

— Veracity[8] (accuracy, reliability of data)

Data and findings extracted from subjective comments and opinions are, by nature, difficult to predict. It is the same for weather data, seismology measurements and macroeconomic indicators. However, this kind of forecasting data is of critical importance to the complex modelling of business scenarios. But despite the relevance of this data, the uncertainty – analytically speaking – cannot be eliminated through just any adjustment methods.

As is so often the case with young technologies, the analysis of the potential of big data is still in its infancy. Several questions remain unanswered. It is possible, of course, to mine the data to identify risks, which can be categorised using theoretical decision-making models and different probabilities. But events may also occur that could never have been predicted, known as black swans.[9] The need to identify uncertainties and to realistically factor them into planning is a further characteristic of data and an ever-present challenge for which not even big data has a silver bullet.

Sources/drivers of big data: The digital revolution across all sectors

The wealth of potential data sources is seemingly inexhaustible. They can be divided into three rough groups based on their origin. Often the data is generated by a combination of these three groups:[10]

— *Machine-generated data*: e.g. sensor or log data[11], language/audio/video, click statistics, data services.

— *User-generated data*: e.g. social networking, correspondence, publications/patents, images, free text, forms, logs, open data/web content, language/audio/video;

— *Business data*: e.g. master/case data, CRM[12] data and transaction data.

The dominant drivers undoubtedly include the ubiquitous digitisation (digital revolution) of infrastructures in the fields of energy, health, transport, education and public administration as well as the increasing personal and work-related use of mobile, internet-enabled devices. It is now possible, for example, to monitor users' internet surfing behaviour around the clock and click by click with the help of tracking software or to determine a person's exact location or patterns of movement using RFID[13], infrared or wireless technologies such as near field communication (NFC). Film and music files are continuously being downloaded and uploaded or streamed in real time. At the same time, millions of users are interacting on social networks and generating digital or viral content that is also disseminated in real time. All this is leading to an exploding volume of data that is potentially valuable but also places huge demands on the infrastructure and architecture of the IT systems.

Mainly internal sources — 14

% of respondents (n per data point=557-867 out of 1,144), 2012

Transactions	88
Log data	73
Event data	59
E-mails	57
Social media	43
Sensors	42
External data feeds	42
RFID scans or POS data	41
Free-form text	41
Geospatial data	40
Audio data	38
Still images/videos	34

Sources: IBM Institute for Business Value, Saïd Business School Oxford

Smartphone data traffic — 15

Petabytes per month

1,263
1,585
4,934
1,803
839

- Latin America
- North America
- Asia-Pacific
- Central and Eastern Europe
- Western Europe

Source: Ericsson 2013

---

[8] Schroeck, M. et al. (2012). Analytics: Big Data in der Praxis. Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. (Analytics: practical uses of big data. How innovative companies are drawing benefit from their data.) Page 4 et seq.

[9] Reference to 'The Black Swan: The Impact of the Highly Improbable' by Nassim Nicholas Taleb (2007). The basic idea is that unprecedented events are by their very nature impossible to predict, whereas risk management deals with situations and events where probability can be calculated.

[10] Cf. Fraunhofer IAIS (2012). Big Data – Vorsprung durch Wissen. (Big Data – Getting ahead through knowledge.) Page 20.

[11] A log file or log is an automatically updated file that contains a record of events or processes generated on a computer system.

[12] CRM = Customer Relationship Management.

[13] Radio frequency identification (RFID) enables objects, animals and people to be automatically identified and tracked and provides a highly efficient means of gathering data.
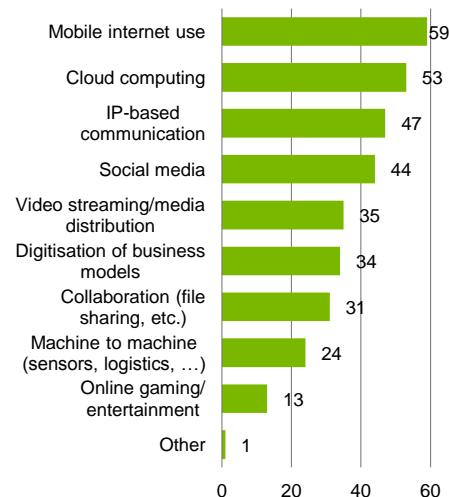
It appears to be the case, however, that social networks do not generate the biggest data sets. For example, the sum of all messages posted on a microblogging platform (e.g. tweets) about specific subjects is not nearly as extensive as the data sets produced by industry and/or academia. Examples of really large data sets include geophysical measurements taken on oil platforms or from seismographs and weather stations, and complex calculations of scenarios in the natural sciences or engineering disciplines. The globalised financial sector also produces a huge amount of analysable data. Every second, its digital trading systems process millions of transactions that can be (and are) analysed in real time.

As the main driver of the big data movement, the digital revolution offers enormous scope for experimentation and lucrative applications, particularly for the key technologies of biometrics and sensors[14]. These technologies act almost like a turbocharger for big data, because sensors in particular can be incorporated into a growing number of everyday (mass-market) objects and machines. People are getting closer to achieving their goal of making everything around them measurable. We will be taking a closer look at this dynamic driver in the following chapter:

3. Sensors and biometrics take over the mass market

Nowadays, every smartphone/tablet or other mobile, internet-enabled device contains a host of different sensors, which goes a long way to explaining why sensors and biometric identification technology are taking over the mass market. Thanks to these devices, the use of sensors and the tracking of individual body metrics are becoming part of daily life. Mobile-phone networks that are becoming faster and more stable are enabling huge numbers of people to be permanently connected to the internet through their mobile devices, which gives additional impetus to the connectivity between 'man and machine' and between 'machine and machine'. The inbuilt sensors give internet services lots of scope to develop new functions, often in the form of apps, and are used millions of times a day.

Temperature, distance, length and depth, light, time, speed, body metrics and weight are just some of the things that can be measured. Nowadays, no field of modern technology involved in measuring, monitoring, automation or control engineering can get by without sensors. This development is being further driven by cutting-edge microelectronics and the miniaturisation of sensors, making them increasingly smaller, cheaper and more powerful.

Mobile devices offer a wealth of sensors …

Mobile devices contain everything from motion sensors, light sensors, altitude meters and digital compasses to fingerprint sensors, gait sensors, voice recognition functions and proximity sensors. The last of these, for example, ensure that your device's touchscreen is automatically deactivated when held against your ear. Biometric identification technology is another growing area of experimentation in the mobile devices sector. Using a motion sensor or microphone, it's possible to measure how you walk, run, speak or drive. Special algorithms are then used to create gait, voice or driving profiles that are stored on the particular mobile device, enabling, for example, smartphones to identify their owners from the way in which they walk, talk or drive.[15]

… and increasingly also biometric identification technology

Sensor technology and biometrics therefore play a vital role in gathering information from our environment and surroundings, from medical processes

Global drivers for big data   16

Survey of German companies > 500 employees, multiple answers permitted [n=100], %, 2012

Mobile internet use	59
Cloud computing	53
IP-based communication	47
Social media	44
Video streaming/media distribution	35
Digitisation of business models	34
Collaboration (file sharing, etc.)	31
Machine to machine (sensors, logistics, …)	24
Online gaming/ entertainment	13
Other	1

Source: A study by the Experton Group AG for BT GmbH & Co. oHG

---

[14] Sensor technology refers to the development and application of sensors for measuring and monitoring changes in environmental, biological or technological systems.

[15] This technology, which identifies people by the unique way in which they move, would allow smartphone owners to authenticate themselves online relatively easily for a wide range of transactions or activities without having to enter passwords or codes. See also Dapp, T. (2012). Homo biometricus: Biometric recognition systems and mobile internet services. Page 7 et seq.

Worldwide mobile device subscriptions 17

Million

- Mobile PCs/routers/tablets
- Smartphones
- Trad. mobile phones
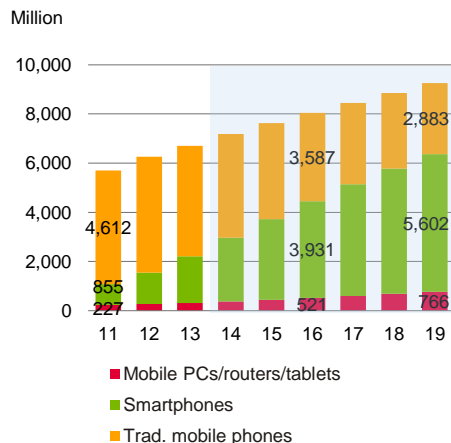
Soource: Ericsson 2013

and procedures, from robotics and automotive technology and from home appliances and office equipment. The constantly growing flood of data that results from this 'smart connection' of objects will go beyond what many people ever thought possible and bring with it new possibilities for connectedness. The volumes of data generated in this way can only be processed using specialised hardware and software. Both for industry and academia, these technologies offer lots of scope for experimentation, research and development, coupled with excellent growth prospects.

All the individual data points are placing new and above all growing demands on existing IT network architectures, whose job it is to process the rising volumes of data quickly, flexibly and at a low cost. A lucrative future is in store for those decision-makers who not only collect and store this data but also convert it into machine-readable structures, identify patterns and draw the right conclusions.

In order to achieve the desired increases in efficiency, however, all data sources must first be a) brought into a machine-readable format and b) collated and combined. Only in this way can the various types of data be correlated with, say, locational data from smartphones or RFID technologies, transaction data from online trading or profile data from social networks. It will then be possible to filter out unexpected, hidden and surprising causal connections that will create value for the respective stakeholders.

The convergence of information and communication technologies as a basis for unexpected innovations

In the future, it will be increasingly common for everyday objects to be connected via the internet so that they can communicate with each other and offer an even higher level of convenience than before. These technologies are being incorporated as flexibly as possible into people's environments and daily lives, and, as a cutting-edge interface between man and machine, are opening up the possibility for new (digital) business models. The merging of information and communication technologies goes hand in hand with the discussion surrounding the 'Internet of Things'. The vision behind this term is of the internet as a bridge between the real and the virtual world, and as a means of making all kinds of everyday objects from people's work and home lives 'smart' (smart home, smart city, smart glass, smart grid, smart car, smart everything). In addition to smartphones and tablets, this could include objects ranging from home electronics, vehicles, traffic lights, parking meters, sports equipment and clothing to entire buildings, delivery containers and industrial machines (known as 'industry 4.0').

Intelligent objects are programmable, equipped with sensors and able to store and communicate information autonomously and often wirelessly using the web. This connectedness can be used to trigger a range of actions or to have one object controlling another. These web-enabled objects could also function as physical access points to various internet services, which would meet the increasing desire of many people for digital mobility. This relatively young area of research holds a great deal of potential that can be unlocked in all kinds of unexpected ways, as the following examples show:

A washing machine that can communicate with garments

In the home, for example, you might soon be able to use your smartphone to remotely control web-enabled blinds, networked radiators or smart air-conditioning systems. There will be washing machines that wait until electricity is at its cheapest before starting their wash cycle. Inbuilt thermostats will be able to make micro-adjustments to the temperature of the water to get the best out of

the detergent's enzymes. Intelligent garments will be able to tell the washing machine that they should be washed only at 40 degrees and on a gentle cycle and that they are blue.

Greater independence in later years

Sensors and cognitive assistance systems could soon be helping older people to live independently and for as long as possible in their own homes, by offering active support in the areas of continuing mobility, physical aid, communication, medical monitoring and medical treatment.[16] It's possible, for example, to install motion sensors into the floor that automatically trigger an alarm when a person falls. Round-the-clock monitoring systems could keep track of data that is relevant to an older person's state of health, such as blood pressure, heart rate, blood oxygen levels and stress levels. Miniature sensors attached to a bathroom mirror could be used to remind people suffering from dementia to take their medication or brush their teeth. With the aid of sensors, it might be possible to measure if a person driving a car breaks out in a sweat or shows other symptoms that might indicate the onset of a heart attack or other incapacitating condition. Technologies could then be developed that respond to these patterns, for example, by slowing down the vehicle, safely bringing it to a stop on the side of the road (using autopilot) and automatically alerting the emergency services.

Potential for connectedness in industry

There is a similar potential for connectedness in the production chains and machinery of a range of industries. In the automotive sector, for example, companies are working on 'networked driving', in which the vehicle's electronic systems are hooked up to the internet. The mobility requirements of car owners are changing. People are starting to expect their vehicles to offer a similar level of comfort and functionality as their web-enabled mobile devices. Soon it won't just be the quality, safety and reliable performance of the vehicle that counts; people will want their digital lives to continue when they get behind the wheel. At the forefront of this development are several trends, including the vision of automated driving, which looks set to become a reality in the medium term, and the long-term move towards more fuel efficient, low-emission vehicles. As the industrial and digital worlds converge, the market will be taken over by new business models that focus on high-tech and primarily personalised services for the driver. These include cloud-based voice recognition, real-time sharing of traffic flow data and proactive, part-automated driving that uses data from online sources and navigation systems to automatically apply the brakes, accelerate and even steer.

Everyday objects with their own web address

As more and more everyday objects get hooked up to the internet for remote contact and control, we will need to look beyond the unilateral communication capabilities of RFID if we are to fully exploit the convergence opportunities. This is why the objects will be given their own IP address so that they can communicate with other smart objects and network nodes. Because of the vast number of IP addresses that this would require, providers will have to use the newer 128-bit IPv6[17] addresses that are gradually replacing the IPv4 versions. However, this also means that further efforts will have to be made to improve infrastructure and expand networks.

Warnings from data privacy experts

Despite all the potential benefits, data privacy experts are very critical in their assessment of the 'Internet of Things', because in many cases it may be

---

[16] Cf.: Beck, S., M. Grzegorzek et. al. (2013). Mit Robotern gegen den Pflegenotstand. (Using robots to alleviate the care crisis).

[17] http://en.wikipedia.org/wiki/IPv6.

possible for sensors to send data to various interested parties for analysis without the individual's consent. Even today, the right of individuals to determine what happens to their information is already being violated.

4. On the role of digital ecosystems in the big data world

Big data is discussed and interpreted in different ways depending on how it is used. It does scant justice to the term if we reduce it to commercial interests or the opportunities presented by social networking applications in digital ecosystems such as Amazon, Twitter, Facebook, Apple and Google.

The positive side to big data

After all, big data also has potential to solve or at least mitigate some of the problems that face society, including in the areas of public health, climate change and international cybercrime. It could also herald a sea change in scientific methods. Whereas researchers have up to now been using preconceived models, such as sample analyses, to produce their findings, they might soon also be able to draw on data-driven observations in real time that use a far larger total sample at no great additional cost. The IT security and financial sectors also stand to benefit: new ways are emerging of bringing credit card fraudsters[18] to justice, for example, and of better assessing risks and compliance; manufacturing companies will be able run their production areas with greater efficiency; logistics chains will be optimised and sensitive infrastructures made more secure. Findings from medical research will be combined with findings from other fields of research as a means of revealing any hidden correlations and/or treating rare diseases.
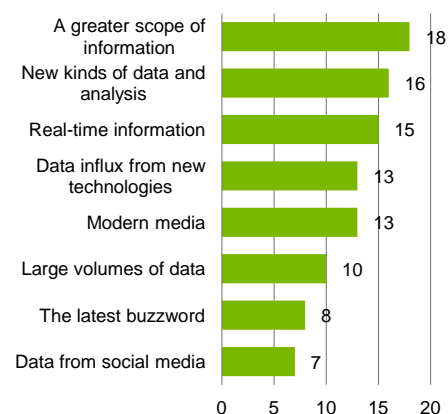
In addition, costs will be continuously brought down, the basis for making decisions will be optimised and it will be possible to more precisely analyse consumer behaviour and customer satisfaction and/or to create products and services targeted at specific types of people.

The dark side to big data

Big data also has a dark side. In the big data debate, people often bring up Orwellian 'Big Brother' methods as imagined in Nineteen Eighty-Four[19] i.e. the excessive use of surveillance. In German we talk of the *gläserner Mensch*, the 'see-through person'. Unlike in Orwell's world, however, the dominant fear in the big data discussion is not that of a dictatorial system. Indeed people will initially participate voluntarily, while in the background their personal information is shared, traded and commoditised by businesses, scientific institutions, NGOs and intelligence agencies. This may sound far-fetched, but there is an understandable concern that citizens and consumers are increasingly becoming a plaything for various stakeholders or are unwittingly becoming such 'see-through people'. In the light of recent revelations about the activities of the intelligence services, it would seem to some observers as if Orwell's novel is closer to reality than ever before.

The objectives pursued by digital ecosystems are more wide-ranging than many assume

On the one hand we have the many useful products and services that help people perform everyday tasks and routine actions, including the aforementioned examples of sensors benefiting the care sector. On the other hand, however, digital ecosystems are accused of engaging in business practices that do not exclusively serve the primary benefit of the product or service for the customer. Once data has been collected, it can be monetised or

Broad interpretation of what big data means `18`

% of respondents* (n=1,144), 2012

A greater scope of information	18
New kinds of data and analysis	16
Real-time information	15
Data influx from new technologies	13
Modern media	13
Large volumes of data	10
The latest buzzword	8
Data from social media	7

Sources: IBM Institute for Business Value, Saïd Business School Oxford

---

[18] Using a multitude of data relevant to credit card crime, the patterns of behaviour common to credit card fraudsters are filtered out and compared with current cases. The aim is to flag up credit card transactions that have a high likelihood of being fraudulent.

[19] George Orwell's novel (1949) is often cited in discussions about misuse of state surveillance measures.

used again and again for sometimes dubious purposes – without the person concerned ever being informed.

Digital ecosystems are giving us 'smart everything'

Graphic: Oliver Ullmann. Deutsche Bank Research.

Source: Dapp, T. (2014). Big Data – The untamed force. Deutsche Bank Research. Frankfurt am Main.

A host of monetisation strategies

If, for example, a company like Google began to offer web-enabled glasses or to conduct research into the development of self-driving vehicles, then it's possible that there might be more to this than meets the eye. Both products could enable virtually limitless access to personal and sometimes intimate data. A single company would gain insights into people's day-to-day lives, their activities, tendencies and identities. Web-enabled glasses could reveal, for example, what people are looking at, for how long and how often their gaze falls on objects, advertising and other people. This would give the company the opportunity to use big data analysis tools in real time as a means of identifying meaningful patterns from the aggregated data, which could then be monetised or used to send people personalised advertising messages. The desire shown by the major internet companies to re-use this additional information again and again for different purposes is growing, while the social consequences are, as yet, largely unexplored.

Automated driving: once a vision, now a reality

The owners of self-driving cars may also unwittingly reveal more about themselves than they intend.[20] Throughout the journey and even before and

---

[20] Morozov, E. (2013). Big Data. Warum man das Silicon Valley hassen darf. (Why you're allowed to hate Silicon Valley.) Frankfurter Allgemeine Zeitung, 10 November 2013.

afterwards, a whole host of information could be measured and gathered for subsequent analysis. After a while, of course, patterns will be observed that will provide answers to the following questions: where do people go when they are driving? At what times of day and night do they use their vehicles? Where do they work and where do they like to go in their free time? How long do they spend in the car on average, what music do they listen to, what temperature do they like the interior to be at, who are they calling or otherwise communicating with, and what digital content are they consuming during the journey? Are they keeping to the speed limit, are they perhaps eating, drinking or smoking during the journey? A company could potentially monetise every activity described here. It could earn this money either by offering tailored products and services from a single source or by selling the personal data on to stakeholders from a range of sectors. Most importantly of all, it would make it possible to track a person wherever they go, thereby bridging the crucial gap between the virtual and the real world.

Revenue does not necessarily have to flow from the end consumer of a particular product or service directly to the provider. Instead, this provider can maximise its income by making the data that it collects during the course of the transaction available to third parties. For start-ups or niche providers it will be possible to latch onto ecosystems with high market shares in order to offer complementary products and services that make the original offering even more attractive for the customer. Takeovers by digital ecosystems are also conceivable, of course.[21] Individual companies would expand their market position and there would be less competition.

Silicon Valley as a global driver of innovation

The cross-sector onward march of …

The digital footprints of internet users are becoming clearer, more precise and easier to track. On platforms provided by popular digital ecosystems such as Google, Twitter, Facebook, LinkedIn, Amazon and Apple, people think nothing of revealing personal and often intimate information such as their relationship status, hobbies, travel habits, consumer habits, favourite music and film genres and photographs, and they'll 'like' virtually anything that their fingers or cursors come across.

Any short message that is followed or created on microblogging services indicates an individual preference. It is relatively easy to create a profile of an individual's political and social views from a list of all their tweets. On Facebook or similar social networks it is relatively easy to find out where people spend their time and what their interests, hobbies and even sexual preferences are simply by looking at what they post and whom they are friends with.

… the online ecoystems[22]

These days, it's virtually impossible to avoid certain digital ecosystems, regardless of whether you're using the internet privately or for work. Practical and popular online services function as one-stop-shops for all possible aspects of life and work and include search engines, operating systems for various end devices, e-mail services, navigation, cloud, streaming and wiki services and proprietary marketplaces for digital goods (Apple Store, Playstore, etc.). People's search queries can be collated and analysed to create profiles even if they are not signed in to the search engine or are not using an e-mail account. The idea is to digitally organise as much information (e.g. personal data) as possible. However, the data is much more useful if it can be assigned to a particular account.

---

[21] See, for example http://www.spiegel.de/wirtschaft/unternehmen/google-kauft-nest-labs-fuer-3-2-milliarden-dollar-a-943362.html.

[22] Cf.: Bahr, F. et al. (2012). Schönes neues Internet? Chancen und Risiken für Innovation in digitalen Ökosystemen. (Great new internet? Opportunities and risks for innovation in digital ecosystems.) Page 3 et seq.

Should you follow the advice to be economical with the data you divulge, you'll find this is only a partial solution – mainly because even 'non-participation' in the digital world leaves behind a trail that can enable conclusions to be drawn about individual people and their behaviour (e.g. by means of de-anonymisation). Anonymising personal data before providing it to third parties does not go far enough to protect privacy, because even individual searches (combined with additional data) can be used to precisely identify the originator of the data – even if their user name or IP address has been deleted, as was shown during the AOL search data leak.[23]

The web as a playground for fenced-in digital ecosystems

The lock-in effect

Google and other internet companies such as Apple, Facebook and Amazon dazzle consumers by offering a whole host of modern, social and often free web-based services. These ecosystems have literally hundreds of millions of loyal customers and enough liquid funds to continuously drive innovation. They heighten their appeal by ensuring a high availability of additional online services and apps within their fenced-in playgrounds. New, bolt-on services appear at regular intervals thanks to strategic alliances with third-party or niche providers or through takeovers. The innovation rate of these technology-driven internet companies is remarkable and is without parallel in both Germany and Europe.

It is also evident that the major online platforms are increasingly extending their feelers beyond their core area of business (which is in any case growing), and in doing so are shaking up existing markets and putting established companies under pressure. Amazon, for example, is investing in its own mobile payment systems as well as its own TV productions and cloud services for premium customers[24], while Google is investing heavily in home appliances and robotics and has even set up a new division aimed at developing humanoid robots.[25] Digital ecosystems are and remain popular and accepted platforms. People are often 'locked-in' to their services, and switching can prove relatively expensive. These walled-garden strategies are giving the digital ecosystems increasing scope to monetise the individual attention and personal data of their users or to use this data for other purposes.[26]

Every minute on the internet …	19
100 hours of video material are uploaded to YouTube	
695,000,000 Google searches are made	
3,300,000 posts are shared on Facebook	
347,000 tweets are sent using Twitter	
48,000 Apple apps are downloaded	
38,200 photos are shared on Instagram	

Sources: YouTube, Google, Facebook, US Securities and Exchange Commission, Apple, Instagram

The ambivalent behaviour of the internet user

Despite all the justified criticism of digital ecosystems, the reality remains that many internet users have (so far) been willing to give up some degree of control over their personal data to the operators of online platforms in order to benefit from their one-stop-shop services. The pull of free, convenient services is obviously so great that huge numbers of people are willing to divulge personal and intimate data without a second thought, are allowing this to be passed on to third parties, are accepting personalised advertising messages and are allowing biometric scans of their voice and face to be taken and quietly stored.

Manipulation and lack of transparency

The users of such platforms have little control over the security of the system being offered, over potential access to their personal data, and over the security, use and deletion of their data. Despite this, the vast majority of people relinquish control of their data to the major internet platforms and assume that the operators will provide the necessary security and will protect against misuse

---

[23] http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=0.
[24] Example: Amazon enters the market for mobile payment solutions:
http://www.businessinsider.com/report-amazon-has-bought-square-competitor-gopago-2013-12.
[25] http://www.zeit.de/digital/2013-12/google-roboterfirma and
http://www.zeit.de/wirtschaft/unternehmen/2013-12/google-roboter-android-erfinder-andy-rubin.
[26] Cf.: Dapp, T. (2013). The future of (mobile) payments: New (online) players competing with banks. Page 21 et seq.

of data. There is a distinct ambivalence to the way in which people behave on digital channels. Critics say the suspicion is being confirmed that it is the internet users themselves who are becoming the commodity to be traded.

For many companies this means first and foremost a lucrative line of business. Most consumers will continue to use the wide variety of personalised services despite possible concerns about data protection. Some will gradually come to the realisation that they are being manipulated and made 'transparent', or will even feel robbed of their data sovereignty. It remains to be seen whether, in the medium- and long term, people will continue to unquestioningly pay the relatively high price of these individualised services (comprehensive, personal, digital profile) or whether they will try to curb their demand for services and products or switch to alternatives. Because of the current situation regarding competition in the market, there is a justifiable question as to whether, in the medium term, serious alternatives to certain online services will or even could become established on the market and whether the speed at which modern technologies are adopted will decline because of the practices.

What do people use their mobile devices for? `20`

54%	surf the internet
37%	spend time on social networks
17%	shop online
21%	share documents, videos, etc.
26%	watch videos and listen to music
16%	make mobile payments

n=4,500 from nine selected countries (DK, FR, DE, IT, NL, PL, RU, SE, GB)
Source: Norton, 2012

The imbalance in data sovereignty

Data protection experts are unanimous in their concern about individual internet companies sharing personal data that is often private and intimate in nature. People are beginning to question whether the fenced-in playgrounds of the digital ecosystems are controlling innovation, communication and information flows and undermining data protection regulations and user autonomy. Moreover, data protection experts are noting that there is a lack of transparency about the ways in which user activities are being monetised and misused.[27]

5. The economic value of data

We have a 'digital twin'

"Many do not realise, or simply do not want to know that they are complicit in the creation of the virtual twin to their real life self – their alter ego who reveals, or could reveal, both their strengths and weaknesses, who could disclose their failures or deficiencies, or who could even divulge sensitive information about illnesses. Who makes the individual more transparent, readily analysed and easily manipulated by agencies, politics, commerce and the labour market"
[German Federal President Joachim Gauck, 3 October 2013, Stuttgart]

The combination of different data sets promises new insights

Individual data sets held by individual companies and viewed in isolation are only of limited use when it comes to analysing big data. Only once several data points, some from different sources, are merged does it become possible to extract certain patterns. Very few companies offer such a diverse range of products and services that they can accumulate a sufficiently broad variety of customer information under one roof.[28] If this information contains enough personal data, it can be used to create comprehensive and relatively detailed profiles of people. The majority of the various data being stored, archived and analysed originates from digital advertising, information, transaction and other web channels. It is relatively easy to collect information about the respective internet provider, IP or e-mail address and the search engine used. With every visit to the Amazon platform, for example, this information is stored and analysed, in combination with each person's historical click behaviour and consumption habits. This is how Amazon can offer its customers personalised purchase suggestions every time they log in, even from different IP addresses, and this is also how personalisation on the internet works in general.

The growing data collections held by the stakeholders makes it relatively easy to link the real and the virtual world, making even physical addresses and

---

[27] See chapter 6 (Walled-garden strategies) of Dapp, T. (2013). The future of (mobile) payments. Page 21 et seq.

[28] Google is very much the exception here, as the company is able to combine its own data, thanks to its reach and its extraordinarily high data volumes.

telephone numbers of individuals fairly simple to identify. On top of that it is possible to make further deductions, e.g. about rents or house prices in a particular residential area, about employers and therefore about salary (at least within the general salary ranges for the industry or sector) and about disposable income. What would happen if personal data was also linked to other information from social networks, data on consumption habits, location data (GPS tracking), data on relationship status and income? The resulting data correlations would in turn allow new, relevant conclusions to be made about personal preferences and characteristics.

How digital footprints become marketable and sought-after profiles

The high price of personal data

Imagine the following scenario: due to some data error, or thanks to a whistleblower, you happen to come into possession of a data set about yourself. Your profile contains quite detailed information originating from a company specialising in data collection or a 'friendly' intelligence service. In your profile you read the following about yourself:

Big Brother is watching you!

"X is male, age 43, an engineer with a PhD, married to V, female, age 37; X lives with person V in a detached house in Frankfurt and has a cat; X subscribes to 'Ideal Home' magazine and also has a digital subscription to the magazine 'The Engineer', which he downloads directly to his brand A mobile device; X regularly buys technical literature, clothes and electronics from online retailer A; X is interested in various environmental issues and supports a number of international NGOs by standing order; X interacts with a number of social networks several times a day on mobile and stationary devices, and maintains an above-average amount of contact with the following persons: H (male, age 41), S (female, age 29), and K (female, age 35); X consumes video-on-demand services and surfs the internet for one to two hours a day on average from 10 pm onwards; X holds several bank accounts with the German banks D and P, as well as one with Swiss bank U; X prefers digital, web-based payment methods; X holds two credit cards from providers M and A, credit card M is also used by person V (wife), transactions for the current month for credit card A: Paris travel guide EUR 9.95, flights to Paris for two people EUR 379.00, hotel in Paris EUR 440.00, restaurant in Paris EUR 90.00, jewellery in Paris EUR 399.00, transactions for the current month for joint credit card M: no transactions; X regularly buys sports goods for outdoor activities; X is increasingly soliciting e-mail offers from various car dealers (i.e. is in the process of buying/leasing/renting a vehicle); X is comparing prices of electricity providers (may be considering switching); X always books an annual winter holiday in Austria online and hires his ski equipment online in advance from two providers (otherwise tends towards long-haul trips with person V and city breaks, the last with person S (X is very probably having an affair with person S)[29]; X obtains prescription medication several times a quarter (X suffers from allergies and high blood pressure), X regularly visits different forums and participates in discussions about rare medical conditions (X uses the anonymous avatars 'snoopy' or 'curious_1970' for this). X holds car insurance, legal insurance and home contents insurance from company A; In 2012 X obtained legal services from company A after being charged with a traffic offence, the case is ongoing. Additional real-time information: person X is at this very moment located at Infidelity Street 7. This is the residence of person S."

Given the range of digital footprints that we leave on the internet nowadays (knowingly and unknowingly), and the data analysis tools required, this type of personal profile is definitely realistic as well as technologically feasible. This

Fear of data theft — 21

Survey, in %
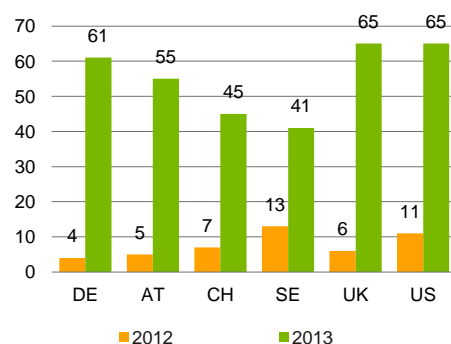
DE (n=571); AT (n=586); CH (n=476); SE (n=346); UK (n=435); US (n=409); Multiple answers permitted

Source: eGovernment MONITOR 2013

---

[29] At this point I refer you back to the previously mentioned question of whether, based on individual transactions, credit card providers can predict which of their customers may be about to face a marital crisis. The answer is: 'Yes, they can'.
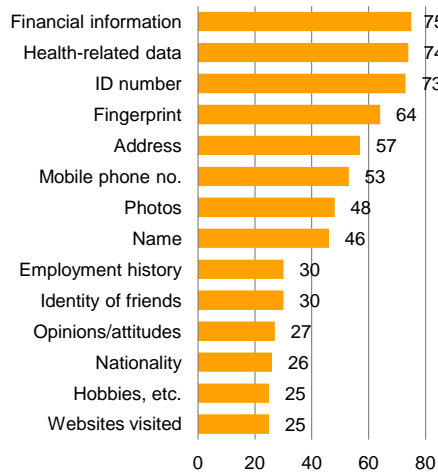
What is personal data?   22

% of respondents, EU-27, 2011 (n=26,574)

Category	Value
Financial information	75
Health-related data	74
ID number	73
Fingerprint	64
Address	57
Mobile phone no.	53
Photos	48
Name	46
Employment history	30
Identity of friends	30
Opinions/attitudes	27
Nationality	26
Hobbies, etc.	25
Websites visited	25

Source: European Commission

enables stakeholders to create cross-context timelines and digital profiles about locations, social relationships, consumption habits and media usage, health, income, employment, etc.[30][31]

Each individual piece of information in this fictional profile has an economic value, with profitable connection and monetisation points for a range of stakeholders from a business (e.g. insurance companies, manufacturers of consumer goods, various service providers) academic (e.g. neuroscience, sociology, behavioural science) or political background (e.g. tax authorities, public administrators, intelligence services). On the internet, people become the subject of mass surveillance, so-called 'transparent individuals', by virtue of their digital footprints, without necessarily realising it at many stages of their online activities (e.g. when (de)activating their GPS signal).

The sum total of this information is aggregated into individual digital profiles that have a price, based on supply and demand. What might be the monetary value of a single customer profile, which is essentially an immaterial value, and most importantly, how could it be determined? This can be illustrated by a simple calculation: the market capitalisation of Facebook, for example, is currently approx. EUR 70 bn.[32] Facebook is advertising the number of its users as more than 1.2 billion.[33] Based on a simple division, that would make the monetary value of the average Facebook user profile around EUR 58. Making the further assumption that only about two thirds of accounts are actually active, the value increases to around EUR 88. This calculated value could provide the basis for negotiations about the price of individual user profiles with investors or other interested parties. Whether this approximation is actually suitable for calculating a return on investment goes beyond the scope of this study, but it does indicate that the motivation behind big data and the trading of data is primarily based on monetary interests.

Personal data becomes a virtual currency on the web

A data-driven life

People are becoming more and more immersed in a data-driven life. Routine, everyday actions are linked to the latest web-based technologies, making life easier, e.g. through the use of apps (internet services). Every click on an online portal, every voice command to a mobile device and each use of GPS saves time and reduces search costs. These services increase efficiency and convenience in everyday life, and demand for them is undiminished. But these internet services are not really free, in fact they come at a comparatively hefty price. Although it does not cost any money to access many of the popular web-based services, the user is paying by (usually voluntarily) providing individual, personal digital data, a fact that many choose to ignore at the moment they make that decision. It is conceivable that future business models may even offer a monetary incentive to the people who are actually supplying the data, i.e. the internet users themselves.

---

[30] Cf.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.)

[31] The terms 'personalising', 'scoring', 'profiling' and 'tracking' are used for this.

[32] As at 24 January 2014, according to http://www.onvista.de/aktien/Facebook-Aktie-US30303M1027.

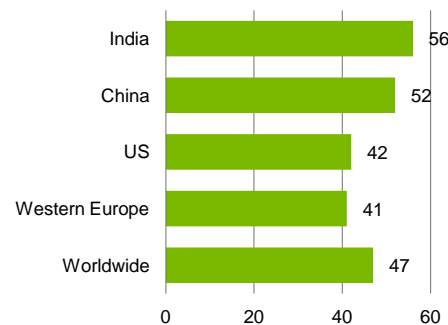[33] http://allfacebook.de/zahlen_fakten/infografik-10-unglaubliche-zahlen-zur-facebook-nutzung-10-jahre-facebook.

Big data – the untamed force

Unprotected data: global problem `23`

%, estimated proportion of all data requiring protection

India 56
China 52
US 42
Western Europe 41
Worldwide 47

Source: IDC Digital Universe Study

The utilisation of personal data in particular attracts many interested parties keen to employ it to a variety of different ends, often with a hope of increasing revenues. Such personal data can be linked directly to an individual. It may be, for example, details of bank accounts or credit ratings, credit card information, information about medical diagnoses and health, data on consumption habits and media usage, data from video surveillance or biometric recognition systems, usage data from applications (apps) or social networks, sensor data from mobile devices, location data and e-mails, chat or other internet communication data. Many people are not (yet) aware of the potential danger posed by a number of different stakeholders and algorithms with a range of interests and objectives collecting and storing digital information in order to analyse it at a later date and potentially monetise it. On the internet, people are being targeted by personalised advertising (e.g. advertising banners) in different places and largely unwittingly, and they are tracked online and spied on with the help of small text files ('cookies'[34]). Once data has been stored it may resurface even years later, and in the worst-case scenario cause permanent damage to people's careers or private relationships. Because it is, of course, possible to manipulate profiles in order to deliberately cause damage to people or discriminate against them, e.g. when they are trying to take out a particular type of insurance, during an application process or through the denial of further services.

As long as there are no solutions that allow for the making of informed decisions, for the refusal to provide private, personal data or for their deletion, the development of this particular aspect of big data with regard to data protection seems far from encouraging.

Personal data is highly desirable

Who controls the much-touted algorithms?

These days, big data is being talked up by many stakeholders from politics, business and academia, for different reasons. Given all the lucrative opportunities for growth offered by big data and by the wide range of modern web-based technologies, the inherent risks and problems are often conveniently ignored. Expectations are high, especially for the new, much-touted algorithms to reduce complexities or to create predictability analyses. But in reality, questions concerning underlying interests, power structures, ethics and morality, monitoring and rights and responsibilities often end up taking a back seat.

The business model of digital ecosystems is based to a large extent on the ability to monetise data. It does not really make any difference who is accessing this personal data, whether it is Google, Facebook or another internet service provider, the police, tax authorities or statistics offices, health insurers, insurance companies, banks, or even – as in the recent controversy – intelligence services. In principle, the use of and access to non-anonymised data by anyone is subject to the provisions of data protection legislation.

6. Big data and data privacy

"IT security is becoming one of the fundamental prerequisites for the preservation of civil rights and liberties. The social opportunities and economic potential offered by digitisation must not be endangered. […] In addition, the subjects of IT security and the defence against industrial espionage should play a special role." That is what is stated in the recent coalition pact between Germany's CDU, CSU and SPD parties entitled 'Deutschlands Zukunft gestalten' (Shaping Germany's future).[35]

---

[34] http://en.wikipedia.org/wiki/HTTP_cookie.
[35] https://www.cdu.de/sites/default/files/media/dokumente/koalitionsvertrag.pdf, P. 139.

Big data – the untamed force

Main problems with big data 24

% of respondents (n=82*), 2012

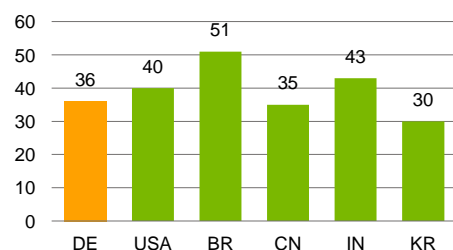|	
Data protection and security	49
Budgeting/setting priorities	45
Technical challenges of data management	38
Expertise	36
Lack of awareness of big data applications and technologies	35
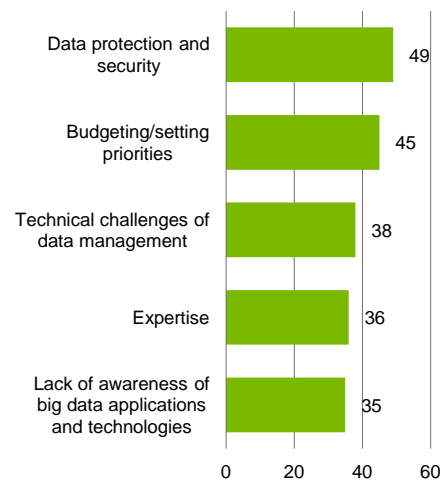
*SMEs and large companies from various sectors

Source: Fraunhofer IAIS

"I try to divulge as little personal information as possible on the internet" 25

% of respondents

DE	USA	BR	CN	IN	KR
36	40	51	35	43	30

DE: n=1,213, USA: n=1,218, BR: n=1,215, CN: n=1,207, IN: n=1,211, KR: n=1,214

Source: Münchner Kreis 2013, www.zukunft-ikt.de

"Companies should be transparent about how they are using personal data and what they are using it for" 26

% of respondents
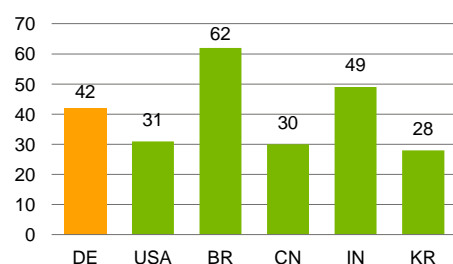
DE	USA	BR	CN	IN	KR
42	31	62	30	49	28

DE: n=1,213, USA: n=1,218, BR: n=1,215, CN: n=1,207, IN: n=1,211, KR: n=1,214

Source: Münchner Kreis 2013, www.zukunft-ikt.de

An ambitious goal for the current federal government. However, there has not been any concrete action since the controversial revelations about surveillance back in June 2013. In principle, there is data protection legislation at national and European level. Germany has the *Grundrecht auf informationelle Selbstbestimmung* (fundamental right to self-determination in regard to information) and also the *Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme* (fundamental right to the guarantee of the confidentiality and integrity of information technology systems). These are meant to ensure that personal data is not stored, passed on or analysed outside of its original context and without the permission of the person it relates to.

The Federal Constitutional Court passed the first fundamental right to data protection as a result of its 1983 population census decision. More commonly known as the right to information privacy, it states that in principle individuals have the right to decide who holds what information about them, when, and under which circumstances, i.e. who receives what data for which purpose. The provision of personal data should normally take place with the informed consent of the person concerned. It is also subject to the purpose limitation principle, which stipulates that personal data may only be used for the purpose for which it was originally collected. Any infringement of this fundamental right, in particular by government agencies, is only permissible in the public interest and must comply with the principle of proportionality. Furthermore, the concept of data economy for information systems requires that as little data as possible should be collected and processed. "The constitutional jurisdiction also demands technical, organisational and procedural provisions for the protection of the right to information privacy. At the core of the realisation of informational self-determination are the rights of the affected person, i.e. the right to access any data stored about oneself and the right – where necessary – to correction, blocking and deletion. The implementation of the information privacy regulations requires independent supervisory bodies." (Official collection of decisions of the German Constitutional Court (BVerfGE) 65, 1 et seq.)[36]

No suitable basis: A directive from 1995

The fundamental right to the protection of personal data was incorporated into the EU Charter of Fundamental Rights under article 8 in 2009. Under EU law, the right to data privacy must be observed not only by the EU authorities but also by the member states. But reforms are inevitable, as data protection legislation is not being consistently implemented and applied by all member states, and the rather outdated Data Protection Directive (from 1995) cannot keep pace with the speed at which web-based technologies are being adopted or the evolution of the internet in general. In January 2012 the EU Commission submitted a proposal for a general data protection regulation for the EU. It is intended to replace the 1995 Data Protection Directive and has been under discussion by the Council of the European Union and in the European Parliament ever since.[37]

The proposed change envisages that in future the scope of the basic regulation will no longer be exclusively linked to the geographical location where the responsible party is based and where the processing of the data takes place, but also to the question of whether the personal data of people in the EU is affected. This so-called 'market location principle' would ensure that large US

---

[36] Cf.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.)

[37] For details on this see: Schaar, P. (2013). Big Brother und Big Data – Was heißt eigentlich Datenschutz auf Amerikanisch? (Big Brother and Big Data – How do you say 'data protection' in American?). Page 5.

companies that store data, such as digital ecosystems, would no longer be able to claim that they are not subject to European law.[38]

Culture clash

It seems that the US has a different attitude towards issues of data protection legislation. But since there are now millions of European citizens using social networks and online stores run by American internet companies on a daily basis this is increasingly leading to a clash of data protection cultures. This provides plenty of potential for conflict. In view of recent revelations about breaches of data protection law it is imperative that the structural and legislative failures and deficiencies in German, European and also international data protection law are remedied as soon as possible. That is the only way to protect privacy and increase people's trust in the digital world. Greater trust in digital infrastructures would provide an ideal catalyst for innovation and growth, giving a boost to the development of the internet.

Proposed solutions need to apply at international level

Anonymising and aggregating personal data

Some of the big data analysis practices violate fundamental concepts of data protection law, in particular those types of data analysis which relate to specific individuals. This conflict could be defused if the data records were anonymised.[39] The possibility of identifying individuals would have to be ruled out for the period that the data was stored and processed. However, the general danger of re-identification would remain. The more individual characteristics are included in a data record, the greater the risk that with the right additional information it would be possible to draw conclusions about the individual concerned. Another anonymisation measure is data aggregation, i.e. the combination of individual data records into group data sets. Data can be deliberately 'blurred' so that analysis will not return exact figures but rather approximate values. That would significantly impede the retrieval of individual information.[40]

Demand for greater transparency

In addition, measures to ensure transparency in the use and analysis of data records might counteract the lack of trust among internet users and minimise data protection violations. This transparency should cover every step in the analysis process, i.e. data collection, the amalgamation with other data sets, the analysis itself and the subsequent use of the results. If the different stakeholders made their analysis practices more transparent, either through a regulative framework or self-imposed measures, people would be able to make an informed decision about whether they want to give consent for their data to be passed on and/or analysed. This would mitigate the current 'black box' character of big data.

International algorithm agreements offer standardisation and certification

Trade between different countries is normally covered by internationally negotiated agreements. Their aim is to promote world trade and the global economy. Similar to the way in which these trade agreements regulate the international movement of goods, an alliance of countries could ratify international algorithm agreements that would harmonise the trade in personal data and make it subject to certification, for example by external algorithm experts[41]. This could standardise analysis methods and make them transparent. Such algorithm agreements would cover not only the relationship between citizen and state (with regard to the practices of intelligence services) but also the digital ecosystems that operate at international level.

---

[38] Cf.: Schaar, P. (2013). Page 6.
[39] Cf.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.)
[40] Cf.: Weichert, T. (2013). Big Data – eine Herausforderung für den Datenschutz. (Big Data – a challenge for data protection.)
[41] Cf.: Mayer-Schönberger, V. and Cukier, K. (2013).

Based on trust

The change in data volumes has led to a change in the nature of data processing, posing an enormous challenge for the protection of privacy. At the moment, we can only guess at the extent and potential of big data. This makes it difficult to cover all eventualities with legal provisions. The potential will develop fully when the people involved feel able to trust web-based technologies, i.e. their privacy is not violated and their fundamental rights are respected and safeguarded.

Users' trust in digital channels has already been dented

Since the documents from former intelligence analyst Edward Snowden were published in June 2013, the discussion about big data has almost always included the surveillance practices of intelligence agencies. Critics believe that anything which is technically feasible in order to obtain personal data is in fact being done, irrespective of any restrictions imposed by (national) data protection laws.

"If you're looking for the needle in the haystack, you have to have the entire haystack first" [Deputy Attorney General James Cole of the US Department of Justice]

National data protection regulations can be bypassed

The problem is that data protection legislation that applies in Germany can be bypassed using technology. For example, the bulk of German telecommunication traffic is routed through foreign countries (primarily the US), because domestic server capacities appear to be overwhelmed by the sheer volume of data. There are reports that data is being accessed, stored and processed, without officially violating German data protection laws.[42] To what extent large parts of digital communications were deliberately diverted via foreign servers, and whether German telecommunication companies are even able to resist the demands for intelligence services is likely to remain a well-kept secret, like so many details of the surveillance scandal.

Surveillance practices could endanger innovation

Economic damage should not be underestimated

Recent controversial surveillance practices could have far-reaching economic consequences that should not be underestimated. The economic danger lies primarily in the fact that in the medium to long term people might adapt their media usage and consumption behaviour, as well as reducing the speed at which they adopt web-based technologies due to diminishing trust and increasing insecurity. That could lead to a slowdown in the development of web-based technologies, in particular for digital ecosystems, but also for many niche providers and start-ups in the ICT sector, who are already increasingly having to promise their customers secure IT infrastructures and operating systems and to plug potential security gaps.

Secure IT architecture is increasingly important

Even if many people do not feel directly threatened by what they have learned about the interception practices of the intelligence services, the constant media coverage could lead to a certain amount of underlying insecurity. The more people feel like they are under constant surveillance on the internet, the more this has an effect on individual development, freedom, creativity and ultimately also on the innovation and competitiveness of an entire economy. 100% data security is, of course, an illusion, and will remain so. But in future the security of IT infrastructures will become more important for users. Companies that are able to offer reliably secure internet services and technologies should be able to benefit from this.

---

[42] http://www.heise.de/security/meldung/Raetselhafte-Entfuehrungen-im-Internet-2053503.html.
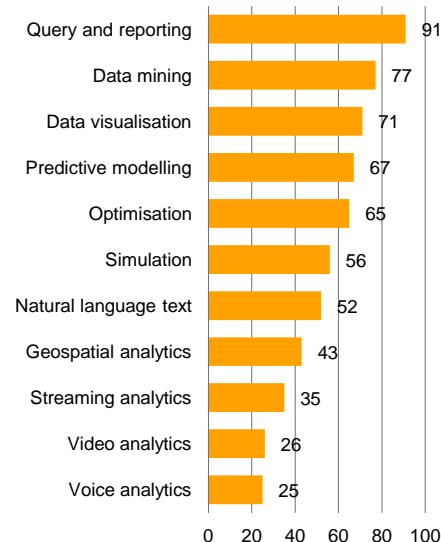
Big data analysis tools `27`

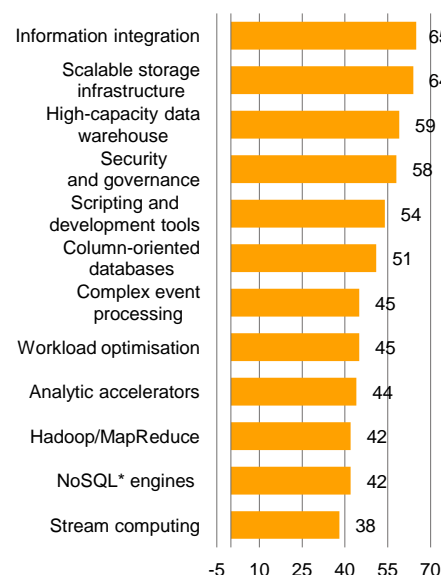% of respondents (n per data point=508-870 out of 1,144), 2012

Tool	Value
Query and reporting	91
Data mining	77
Data visualisation	71
Predictive modelling	67
Optimisation	65
Simulation	56
Natural language text	52
Geospatial analytics	43
Streaming analytics	35
Video analytics	26
Voice analytics	25

Sources: IBM Institute for Business Value, Saïd Business School Oxford

Which platform components are being used? `28`

% of respondents (n per data point=297-351 out of 1,144), 2012

Component	Value
Information integration	65
Scalable storage infrastructure	64
High-capacity data warehouse	59
Security and governance	58
Scripting and development tools	54
Column-oriented databases	51
Complex event processing	45
Workload optimisation	45
Analytic accelerators	44
Hadoop/MapReduce	42
NoSQL* engines	42
Stream computing	38

*NoSQL = Non-relational databases, suitable for data-intensive applications such as the indexing of large document sets

Sources: IBM Institute for Business Value, Saïd Business School Oxford

7. (Big) data in practice: There is no ideal solution

Faced with the potential uses and above all the commercialisation opportunities provided by the explosion of data and the information derived from this, stakeholders need to decide whether it is worth their while to extract and analyse the data that is – in principle – available to them, and whether they should be investing in big data projects generally. The answer is a definite 'yes'. Just as providers are now no longer able to ignore discussions and reviews of their products and services by their (potential) customers on social networks or forums[43], they are also well advised to utilise existing internal and external data sets. Otherwise companies stand to lose valuable competitive advantages due to a lack of relevant informational edge.

The stakeholders would do well to recognise the abundance of information and data streams at their disposal as a potential growth area, and learn to channel it effectively. To this end they require an adaptable digitisation strategy. For example, new IT systems that are able to more quickly detect trends and produce forecasts could be integrated into existing IT architectures, e.g. to provide analyses and forecasts relating to the behavioural patterns of customers, business partners and competitors. The ability to design more efficient processes and structures, optimise IT systems, achieve potential synergies and thereby reduce costs are also benefits that should not be overlooked.

In order to remain competitive, the stakeholders need to learn to combine their growing data pool with the latest management methods as a means of developing suitable business strategies and appropriately adapted products and services. Even if some protagonists may initially be unable to achieve a big breakthrough, many business processes across a range of industries have potential for greater efficiency that can be leveraged in small incremental steps. For financial institutions that could mean, for example, that they pay less attention to financial products (e.g. derivatives) and other new, virtual products, but rather concentrate on the banking services that are being created around the financial sector (such as web-based consultancy, information services, forums). In future, the service portfolio of banks is likely to feature more options based on filtered customer information.

Potential for German SMEs to catch up

According to a survey of innovation potential carried out by the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), respondents stated that they saw cross-sectoral and long-term potential for the individualisation of services (e.g. daily health checks, individual entertainment programmes on demand).[44] They also wanted more intelligent products (e.g. self-learning and self-regulating houses, self-driving vehicles). At the same time, 95% of survey participants voiced their desire for support in the form of best practice examples or training courses, in order to reduce existing knowledge gaps in the area of big data.

At the moment, the great potential offered by the wealth of data available from various sources is being largely ignored by many companies, particularly small and medium-sized ones. That is bound to change in future. When it does, German SMEs will join the ranks of companies facing the enormous challenges of capturing, storing and processing the accruing volumes of data in order to become more efficient and profitable. However, unlike large internet companies, many mid-sized businesses lack the necessary digital strategies and expertise

---

[43] Cf.: Dapp, T. (2011). The digital society. Page 7 et seq.
[44] Cf. Fraunhofer IAIS (2012). Big Data – Vorsprung durch Wissen. (Big Data – Getting ahead through knowledge.)

that would allow them to integrate modern data mining technologies into their internal and external value creation processes. Those who do not respond quickly enough to this development with appropriate and above all flexible software solutions and extraction tools might miss out on valuable competitive advantages and could even lose market share. But the implementation of the latest big data tools is not without its pitfalls and requires additional resources. Irrespective of the different stakeholders involved, the decision-makers should be able to give general answers to the following questions at the start of the big data process:

The four phases of big data adoption		29
1. Educate	Focused on knowledge gathering and market observations	24% *
2. Explore	Developing strategy and roadmap based on business needs and challenges	47%
3. Engage	Piloting big data initiatives to validate value and requirements	22%
4. Execute	Deployed two or more big data initiatives, and continuing to apply advanced analytics	6%

* Percent of all respondents, 2012 (n=1,061, percentage does not equal 100% due to rounding).

Sources: IBM Institute for Business Value, Saïd Business School Oxford

— What are our objectives? For example, increased efficiency and/or revenues, optimisation of IT network and system architecture, promoting open data processes (e.g. implementing external knowledge bases, preparing and publishing data), creating customer profiles, optimising scenario and/or model calculations

— Which data is relevant for the analysis?

— Which data is already available; which additional data could be aggregated, either internally or externally?

— Which analysis technologies and techniques are required?

— Which additional resources are required and who has the authority to decide on them? Expertise (internal/external technical and management skills), internal/external hardware/software (analysis tools, data warehousing, storage capacity), time, financial resources (internal/external financing), etc.

To highlight the potential and the wide scope of possible applications for big data, the following sections are going to give a brief overview of some successful big data projects from business, academia and the public sector:

A big data project in the health sector

Up to now, health authorities have had to rely on the number of reported cases when analysing flu epidemics. But since many people tend not to go to the doctor at the first sign of illness, and the official notification process is fairly slow, there is usually a considerable time lag of one to two weeks[45] before authorities can react to outbreaks, which is rather a long time when it comes to public health.[46]

Big data provides information advantages

A big data project from Google may be able to help. Since the end of 2008 the online tool Google Flu Trends has been available free of charge to users in over 25 countries.[47] A data-based early-warning system for flu epidemics, health authorities can use it in addition to traditional methods for monitoring flu outbreaks to enable them to take timely, possibly even preventive action.[48] Google researchers Jeremy Ginsberg and Matthew Mohebbi developed an algorithm that examines the 50 million most common daily search queries for terms relating to flu. Search volume analyses (also conducted with Google tools) indicate a spike in the number of certain search queries at the start of flu epidemics. Data mining reveals a pattern that shows particular seasonal volatilities. The subsequent comparison of the figures for actual cases, provided by the relevant health authorities, with the predictions based on electronic search query figures shows an almost identical picture. Since there are relatively

[45] Ginsberg, Jeremy et al. (2009). Detecting influenza epidemics using search engine query data.
[46] Cf.: Mayer-Schönberger, V. and Cukier, K. (2013). Big Data. Die Revolution, die unser Leben verändern wird. (Big Data. The revolution that will change our life). P. 7 et seq.
[47] http://www.google.org/flutrends/intl/de/about/how.html.
[48] This model is currently restricted to the analysis of flu epidemics and dengue fever – there are no plans for predicting other epidemics in the near future.

Big data – the untamed force

large regional differences in the spread of influenza, Google uses information about incidents of the disease from regional providers.

Overall this gives a fairly realistic picture without a time lag – as opposed to the traditional flu monitoring methods that tend to have inbuilt delays. On this basis it is possible to produce exact estimates for the current flu situation in different regions. This information advantage buys health authorities precious time by allowing them to recognise flu epidemics sooner and possibly even limit their extent.

Frequency of flu cases in the United States | 30

Historical estimates
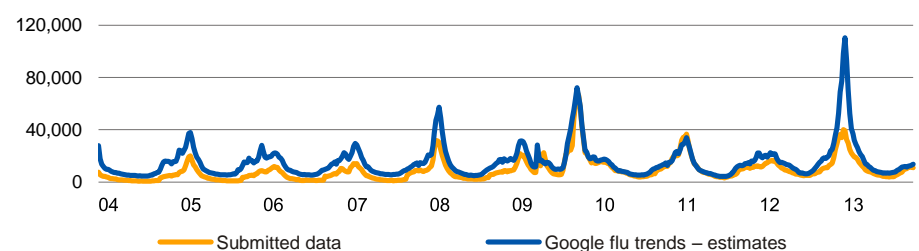
Sources: U.S. Centers for Disease Control; Google flu trends (http://www.google.org/flutrends)

The Google analysis model is reviewed every year and, if necessary, adapted to changes in the electronic search behaviour for health information, e.g. with regard to terminology.[49] The Google predictions for the start of 2013 overestimated the actual illness rates (see chart). Google announced that it was adapting its algorithm to the new search habits. Scientists suspect that media warnings of an impending flu epidemic at the start of 2013 may have led people to 'google' flu-related terms on a purely preventive basis, without suffering any of the actual symptoms. That would have distorted the predictions of the Google tool – at least for the US.

Suspicions remain

Instant flu predictions allow the health sector to be managed more effectively. Google's analysis tool is an example of the potential for innovation offered by big data, along with a possible economic benefit. But we must not forget that to make these predictions it is necessary to store and evaluate personal data. Google assures us that individual users cannot be identified as the analysis of the search queries is anonymised. Only the big picture is of interest, as that is the only way to arrive at reliable conclusions. But is Google complying with data protection legislation? Although the analysis is based on the amalgamation of billions of different search queries from individual users it still needs to use geographical information based on IP addresses in order to be able to make regional predictions.[50] And data based on IP addresses can easily be de-anonymised. In the end, some suspicion always remains. Suspicion that such tools are also an elegant way of justifying the permanent storage and potential monetisation of personal data.

---

[49] In 2009 the original model had to be adapted early on to enable the best possible analysis of the development of the non-seasonal swine flu pandemic. See: Cook, Samantha et al. (2011). Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic.

[50] Internet Protocol address: every device on a network is assigned an address made up of a series of numbers; this allows the accurate addressing and delivery of data packages.

A big data project in the marketing sector

"What scares me about this is that you know more about my customers after three months than I know after 30 years." [Reaction of Lord MacLaurin, former CEO of Tesco, to the initial results of loyalty card trials]

Big data objectives: Customer focus `31`

% of respondents, weighted and aggregated, n=1,067, 2012
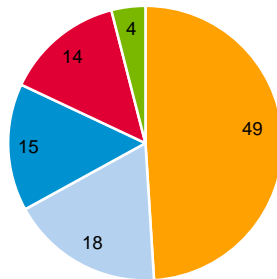
- Customer-centric outcomes
- Operational optimisation
- Risk/financial management
- New business model
- Employee collaboration

Sources: IBM Institute for Business Value, Saïd Business School Oxford

Large (US) retail chains in particular are now commonly using data mining for marketing purposes. Until recently, the relationship between large retailers and individual customers was fairly anonymous. Now 'one-to-one marketing' – based on the model of the village store but on a much larger scale – is intended to transform this into a very personal bond between retailer and customer.[51] Due to his detailed knowledge of his clientele, the proprietor of the village store knew exactly which products to recommend to which customers at which particular time in order to make a sale. Supermarket chains are just beginning to employ the concept of this somewhat analogue 'database' on a massive scale. To this end they are continuously storing, updating and analysing customer-specific data. This data originates mainly from credit card sales and purchases from their online shops – but loyalty cards (customer or bonus cards) also provide information about people's preferences.

Traditional market research with its personal surveys is increasingly being replaced by data-based real time analysis of shopping habits. To find out which group of customers is buying which product thus requires good programmers and data analysts, as well as adaptive algorithms. With accurate analysis results it is possible to offer the customer the right product at the right time, in the form of vouchers, personalised advertising brochures, promotional e-mails, etc. From the retailers' perspective, big data applications primarily have the potential to become a lucrative marketing tool.

The US retail chain Target set out to examine the shopping habits of female customers who signed up for their baby registries.[52] They found the following shopping patterns: from the start of the second trimester onwards pregnant women increasingly bought unscented lotions. A few weeks later they began to buy special food supplements. The changes in their shopping behaviour over the course of the pregnancy showed such consistent patterns that the retailer was even able to predict due dates with relative accuracy. Based on these findings, Target put together product lists that constituted an early warning system for customer pregnancies. If the shopping habits of a female customer suddenly correspond to these 'pregnancy patterns', supermarket chains now tend to target them with specific advertising or vouchers. This personal targeting has become so sophisticated and accurate in its predictions that it can even trigger family crises. In the US a father accused Target of trying to encourage his teenage daughter to become pregnant by targeting her with specific advertising for baby products. As it turned out, the retailer had some inside information. Its analysis of the daughter's shopping habits matched the pattern displayed by many women in their first months of pregnancy. The daughter really was pregnant.

Targeting customers with a personal touch

Over the course of their involvement with big data, Walmart, the world's largest retail chain, have analysed a range of data sets relating to the shopping habits of their customers and discovered some remarkable and, as it turns out, profitable correlations. For example, the snack food Pop-Tarts[53] are bought particularly often if the weather centre has issued a hurricane warning.[54] Walmart also discovered that beer and Pampers are often bought together in

---

[51] Bloching, B., et al. (2012). Data Unser. Wie Kundendaten die Wirtschaft revolutionieren. (The Store's Prayer. How customer data is revolutionising the economy.)
[52] Cf.: Mayer-Schönberger, V. and Cukier, K. (2013).
[53] http://de.wikipedia.org/wiki/Pop-Tart.
[54] http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=0.

the evenings.[55] As a result Walmart has started to use big data marketing tools to optimise its shelf space. Branches now position nappies and beer on adjacent shelves, and when a hurricane approaches pallets of Pop-Tarts are put next to the checkouts.

A big data project in academia

> *"Torture the data long enough, and it will confess to anything."* [Ronald Coase, economist]

The FuturICT Knowledge Accelerator and Crisis-Relief System[56] is a European research project in the field of ICT. This large-scale project, which is supported by a number of international research institutes, was initiated by Dirk Helbing, a professor at ETH Zurich.

The philosophy of open innovation adds another dimension

Its vision is the development of a knowledge platform based on a 'super computer'. It is intended to integrate all global data streams into a model of reality that is as accurate as it can possibly be. Computer simulations and data mining are used to minimise or plug any gaps in knowledge of a technical, commercial, social or ecological nature. One particular aim is to uncover interdependencies between very diverse areas. Virtual simulations of future scenarios help to manage these interdependencies and highlight preventive measures that can be taken to mitigate or even avoid looming crises in world events, e.g. financial crises. To make this possible, FuturICT is to be based on three pillars: a planetary nervous system, a living earth simulator and a global participatory platform.

The objective of the planetary nervous system is the visualisation of current world events, based primarily on data derived from a global network of sensors. This model of the global society is based on reality mining, i.e. the real-time mining of data streams produced in all spheres of life and work. The greatest challenge will be the constant adaptation to new circumstances, because, as the Greek philosopher Heraclitus once aptly said: "The only constant is change", i.e. people and their environment are subject to a permanent process of transformation.

The second pillar, the living earth simulator, is intended to simulate possible future scenarios, based on the data stored in the planetary nervous system and the participatory platform. This political 'wind tunnel'[57] is meant to provide some insight into how the global system would behave as different parameters change. This is where the 'what happens when' scenarios are played out. Each scenario is assigned a probability rating which provides the basis for the development of an early warning system. Factors that could (potentially) destabilise the entire system are, of course, of particular interest. The simulation is based on experimentation with different models and their combination (=pluralistic modelling), embedded in an open-source software environment. This is intended to give a wide variety of researchers the opportunity to directly feed in newly developed analysis tools.

The idea of the interactive infrastructure carries through into the open participation platforms: these are based on the concept of 'prosumers', i.e. consumers who also produce. 'Prosumers' are able to view data streams, simulations and applications that are made available on the platform and also to feed information of their own choice into the system. In this complex system, all users, whether they're individuals, companies or organisations, retain control over their own data and have the opportunity to see the ideas entered by others.

---

[55] http://www.forbes.com/forbes/1998/0406/6107128a.html.
[56] www.futurict.eu.
[57] Geiselberger, H., Moorstedt T. (Hrsg.) (2013). Big Data: Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.) Suhrkamp Verlag. Berlin.

Through this interactive participation, the FuturICT model is given a kind of 'open innovation' component that heavily involves external knowledge owners and thus ensures an ongoing process of quality control. The core concepts of FuturICT should therefore be openness, transparency and participation.

The financing hurdle

When and to what extent FuturICT will be implemented, however, is uncertain after it failed to secure EUR 1 bn worth of funding from the 7th EU Framework Programme for Research and Innovation 'Future Emerging Technologies' (FET)[58]. Despite the potential benefits to society as a whole, there are doubts about the feasibility of FuturICT. A complex global system to which a multitude of users have permanent and simultaneous access requires vast server capacities. Given the current technological possibilities, it is questionable whether a project of this magnitude would be feasible or financially viable – despite the advances in server capacity.

Hayek's criticisms more relevant than ever

Another hurdle is the absence of a credible theory on human behaviour on which a forecasting model could be based. Critics also argue that the world is more complex and chaotic than any model or combination of models could ever be. What we don't know is becoming greater than what we do know, and the compulsion to find a higher degree of coherence than is currently possible quickly leads people to prematurely generalise and simplify complex correlations. In Friedrich August von Hayek's prophetic book 'The Road to Serfdom', social planners, known as the 'social engineers' of mechanism, are attacked for having been taken in by the constructivist delusion that a society can be designed on a drawing board.[59] It also remains doubtful whether we would even be able to understand the recommendations of the living earth simulator. After all, implausible suggestions that are open to interpretation can hardly be implemented as policies.

Open (big) data projects in public administration

Many big data projects are solely driven by the monetisation strategies of particular interests. However, the various data applications, regardless of their data volumes, can also provide a valuable economic benefit to society. Unlike the big data applications that are usually talked about, 'open data'[60] (or 'open government data') does not serve primarily monetary purposes, but it does also offer an often underestimated wealth of data that different stakeholders can use multiple times for any number of reasons. Open data is digital, anonymised information that is made publicly available with almost no restrictions.

Data-driven innovation

Open data is usually focused on non-textual material, such as maps, satellite images, geospatial data and environmental data, rather than documents and records. This movement is driven by the growing demand of citizens for transparency, interaction and collaboration. Another aspect that could present an economic benefit is data-driven innovation. After all, these 'open' data sets can be used freely by journalists, researchers, companies and members of the general public for all kinds of different purposes, both private and commercial, and, most importantly, they can be re-used again and again. This freely available data, in combination with web-based technologies, creates an economic benefit in the form of new business models and innovative products and services. Journalists, for example, are using various open data sets to augment their investigative reports, thereby driving forward the field of 'data journalism'.[61] Open data is there to be re-used and recycled. To support this, the

[58] http://cordis.europa.eu/fp7/ict/programme/fet_en.html.
[59] Hayek, F. A. von (2004). Der Weg zur Knechtschaft (The Road to Serfdom). 4th edition. Mohr Siebeck Verlag. Tübingen.
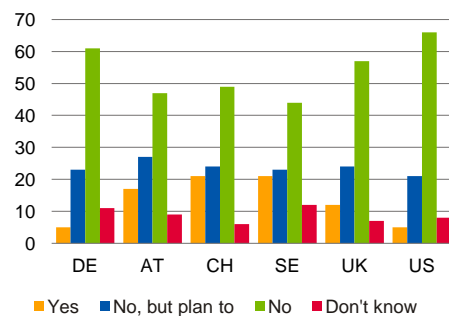[60] http://opendatahandbook.org/de/why-open-data/index.html.
[61] http://datajournalismhandbook.org/.

Big data – the untamed force

Use of e-government services via mobile devices `32`

% of survey respondents, 2013

Yes   No, but plan to   No   Don't know

DE (n=600); AT (n=713); CH (n=702); SE (n=725); UK (n=713); USA (n=576)

Source: eGovernment MONITOR 2013

Establishing standards in open data

Open Knowledge Foundation[62] (OKF) in Germany organises regular 'hackathons' [63] for a wide variety of people including journalists, open-data experts and software developers. They use these events to share ideas and tips about the mining of existing, freely available data and the design of new web-based services in the form of apps.

Although open data can actually cover all types of data, the term is often used as a synonym for open government data. The gradual liberation of government data is changing the relationship between citizen and state. This data that is being made freely available is adding a new dimension to democracy; the previously opaque workings of authorities are becoming more transparent. The US and the UK in particular (data.gov; data.gov.uk) as well as certain Nordic countries are playing a leading role in the publication of government data. Various open data projects have already been successfully implemented in these places and are being used by citizens.[64]

Several years ago, a list of standards was developed that is today known as the *Ten Principles for Opening Up Government Information.* These are there to help governments and administrative bodies make their data records available to the general public. In mid-2013 the G8 member states[65] signed the Open Data Charter, which committed them to five principles [66] that would see them open up their data by 2015. An example of open data is the website that gives the UK's taxpayers a breakdown of what their money is being spent on[67]. This option also exists in Germany: the website offenerhaushalt.de, another OKF project, provides in-depth information about government spending at a national level and includes details about the size of budget available to each department. Since 2005, when Germany's Freedom of Information Act[68] was passed, every citizen has had the right to demand information about such matters. The data on the website is presented in a way that makes it quick and easy for people to access, analyse and process this information.

In the UK, however, it is not only information on how the government spends tax money that is available online. Open data at a local level is being provided to the general public in a similar scale on the website OpenlyLocal[69] Nearly one quarter of the UK's local authorities have made their data publicly available here, and more are being added every week. As well as breakdowns of spending, the data published on the site includes all other municipal information such as population statistics and local financial data.

An organisation based in London (MySociety) has launched a website called 'Fix my Street' that allows people to report potholes, broken traffic lights and street lights directly to the relevant authorities. Thanks to the accompanying app, smartphones can be used to send photos of specific potholes to the relevant authorities, who can instruct repair jobs using the GPS data. The person who reported the fault is given regular updates about the progress of the repair. This concept has since been imitated in a number of European countries. These kinds of platforms can benefit both the municipal authority and the

---

[62] The OKF, for example, has defined a set of ten data sets that national governments should, as a minimum, publish in the form of open data and it rates countries on this basis in its Open Data Census. The most recent census, which now covers almost 80 countries, was conducted for the G8 Summit. See http://census.okfn.org/.

[63] See http://energyhack.de/ and http://jugendhackt.de/.

[64] For an example of this, see the Open Data Institute (http://theido.org).

[65] Germany, France, UK, Italy, Japan, USA, Canada and Russia.

[66] Open Data by Default, Quality and Quantity, Usable by All, Releasing Data for Improved Governance, Releasing Data for Innovation.

[67] http://wheredoesmymoneygo.org/.

[68] The Informationsfreiheitsgesetz gives all citizens the unconditional right to access administrative information from federal authorities. Their wish to do so does not have to be based on a legal, commercial or other interest.

[69] http://openlylocal.com/.

residents – however, there are not always sufficient public funds to repair the fault immediately.

In Germany there is still no central database in which all municipal administrative data is collected. Individual cities, such as Cologne[70] and Frankfurt, are now offering these kind of tools, however. The website 'Frankfurt gestalten – Bürger machen Stadt'[71] (another OKF project), is designed by the people of Frankfurt for the people of Frankfurt, providing information about the city free of charge. In addition, people can use the OpenStreetMap wiki to search for existing data sets relating to local politics and other matters e.g. police reports and roadworks notices. There are also various forums where citizens can discuss local developments.

The projects described in this chapter demonstrate that the focus of open data – as an alternative use of data in the big data discussion – is much more on benefiting society than on earning money. This raises the question as to how much the big data discussion is able to give additional momentum to open data and to the creation of value. The opening up of previously inaccessible data is not only relevant for the public sector, it also undoubtedly has potential for other areas. In education (Open Research Data), for example, we can observe a liberation of data in how departments are making whole series of lectures including the accompanying handouts publicly available on YouTube or on the university websites (e.g. the Massachusetts Institute of Technology's MITOpenCourseWare[72]).

8. The limits of big data

Big Data is not a panacea. Besides the need for data protection legislation to catch up, there are other aspects that highlight the weaknesses and limitations of the latest big data analysis methods. Innovative technologies and processes in the area of big data do open up new ways in which stakeholders can, for example, augment existing data models and scenarios with real time data in order to obtain more meaningful results and ideally forecast trends. However, big data is not able to perform miracles, because the laws of statistics don't simply cease to apply due to a massive volume of new data (correlations). Just because data-driven analyses can suddenly be augmented with new, previously unimaginable quantifiable data sets, this does not make the results any more objective than in the pre-big data era, i.e. despite the increasing ability to quantify human behaviour, and even given all the new, machine-readable buying preferences and expressions of human emotion this does not necessarily produce any reliable facts. This is particularly true of data originating from social networks.[73] No matter how quantifiable and dispassionate the analysis methods, the evaluation of the results from this scientific data analysis still involves interpretation by human beings (with their subjective views), who are fallible.

Increasing complexity in the amalgamation of data sets

Furthermore, there are other methodological problems, such as the unreliability, susceptibility to error and incompleteness of large data sets. These problems are amplified when data from different sources is aggregated (data delineation). When different data sets are merged the complexity of the underlying data pool increases with each correlation. Amalgamation doesn't just mean a simple

[70] http://www.offenedaten-koeln.de/.
[71] http://www.frankfurt-gestalten.de/.
[72] http://ocw.mit.edu/index.htm.
[73] Cf.: Boyd, D., Crawford, K. (2013). Critical questions for Big Data. Provocations for a cultural, technological and cultural phenomenon.

Big data – the untamed force

<table>
<tr><td>Increasing complexity of
data combination</td><td>addition of new data. For example, when ten details from environmental sensors are linked to ten details relating to traffic volume this does not result in 20 new data records, but is instead multiplied to produce 100 new pieces of information, which can all be interpreted in a variety of new ways. This increases the complexity of the newly created data set and the requirements made of it enormously, with the inherent danger of trying to detect patterns in the underlying analysis that do not exist. This could result in the wrong actions being taken.</td></tr>
</table>

Cause and causality

<table>
<tr><td>Big data does not offer
causality, only correlation</td><td>Although data is getting bigger all the time, the maxim of 'quality over quantity' still applies. Big data does not equate to completeness. The available information and data arrays can be examined for correlations using existing analysis tools. But these connections found by algorithms are not always logical. Even with today's advanced technology it is not possible to establish or examine the causality behind the connections found; this task still falls to the human mind, using its experience and intuition. It is down to human beings to determine which questions and answers can be derived from the data. Algorithms are simply an amalgamation tool and determine, *which* connections the data exhibits, but not *why* it exhibits them. Data analysis remains the basis for decision-making, but the power to make decisions should not be transferred to algorithms. Although big data is based on correlations it is unsuitable for assessing their causality.[74]</td></tr>
</table>

What about human intuition?

<table>
<tr><td>The threat of information
misuse</td><td>The new technologies and methods might tempt people to stop trying to find causal connections by using or developing models, and rather let the data speak for itself, without any further interpretation. In the long term that could change the way people think, as they come to rely too much on big data without questioning causes and connections. Thought patterns may potentially be led along certain paths, with a negative impact on human intuition and creativity. But both of these are crucial elements in the process of innovation. Should data algorithms be allowed to automatically initiate recommended actions without further interpretation? The potential preventive use of data evaluation should not be allowed to become common practice. In an extreme scenario, the police might, for example, increase their surveillance of someone simply because his current 'digital footprint' (e.g. tweets, search queries, credit card and online purchases) fits the typical behaviour pattern of known bank robbers as determined by algorithms, without that person ever committing any criminal acts or even contemplating them. At first sight the idea of being able to predict crimes using big data before they are committed does, of course, seem very tempting. But it is in fact a misuse of information if wholesale and unwarranted profiling leads to people getting caught in an electronically created net and put under general suspicion. That sort of behaviour constitutes a general danger to a free and democratic system and was consequently the subject of debate even before the recent surveillance revelations.</td></tr>
</table>

Questionable data

<table>
<tr><td>Data from social networks
does not provide universal
insights into human
behaviour</td><td>Some stakeholders, including academics, like to use data from social networks to try and detect new patterns based on people's moods and feelings in order to explain or even predict media events, political protests or other social movements. But even the analysis of such large data sets does not allow us to</td></tr>
</table>

---

[74] Cf.: Mayer-Schönberger, V. and Cukier, K. (2013).

draw general conclusions about human behaviour. For example, even a large proportion of the Facebook community is still only a specific subset. Depending on the analysis and the message derived, the users of social networks may well differ from other people. A variety of characteristics, such as age, gender, level of education, internet affinity and the willingness to post personal pictures and messages play an important role.

It is impossible to be certain that the officially stated numbers of accounts on social networks do not include fake accounts, or that the number of users and accounts match. Nor can we rule out the possibility that individual social networks censor problematic user content, thereby undermining the value of these messages as a data source from the outset. It is imperative to keep these questions of methodology at the back of one's mind when deciding which questions or what analysis are valid, and consequently which interpretations based on the analysis are reliable.[75] In spite of, or even because of, the big data movement it is still true to say that smart data can provide equally good results.

Ethics and morality

What exactly do we mean by 'open access'?

A comprehensive discussion about ethics and morality within the context of big data would go beyond the scope of this study. But ethical and moral considerations do, of course, play an important role when it comes to the processing of personal data from different sources and contexts, sometimes repeatedly and by different stakeholders. Many questions in the big data discussion remain unanswered, e.g. with regard to access, purpose, control, power structures and veracity. What, for example, is the status of data that is marked as 'public' on the various social networks, and therefore not just open to friends but in effect to all internet users? Should interested parties be allowed to use this data for analysis purposes or is there some sort of 'volunteer protection' which means that the informed consent of each individual user is required? The fact that some digital content is publicly accessible does not automatically mean that anyone can do whatever they like with it. The internet is not a legal vacuum.

9. Conclusion

Big data is unstoppable

Massive data volumes are now a fact of life. Big data is as impossible to stop as globalisation. With the advent of big data, IT security has become a crucial part of the overall value creation process. Where stakeholders used to struggle with limited data, they are now struggling with the limitations of analysis, and the question of which additional valuable information they might be able to coax from the glut of data at their disposal.

Although decision-makers are well aware that big data is a subject of great strategic relevance, offering lucrative opportunities for growth in the medium to long term, they often lack suitable digitisation strategies, trained staff to face the new challenges and the necessary management skills. Inadequate or outdated data protection regulations, coupled with a rather rigid silo mentality, mean that inhibitions about actively experimenting with big data still remain.

New types of jobs created

Big data will alter the structure of existing economic sectors, as new competitive constellations are created. Many established business models are going to be overthrown by the digital revolution, with firms facing increasing competition by companies from other sectors who specialise in products and services relating to web-based information and communication technologies or data analysis. There will also be new types of jobs, such as data analysis specialist and algorithm expert.

---

[75] Cf.: Boyd, D., Crawford, K. (2013).

In particular, start-ups and niche providers who specialise in the evaluation and analysis of various data, including publicly available data, and turn it into products and services, will become much more prevalent. This will increase the pressure on established business models. An application for web-enabled devices that evaluates people's individual media usage and reading behaviour can provide a much better indication of what people like to read, in a much shorter time frame, than any publishing company with years of experience in customer service was ever able to.

Letting data speak

Painstakingly acquired expertise within established business sectors will in future be challenged much more quickly, as the business models of the technology companies entering the market are based on the premise that intelligent algorithms will provide more accurate forecasts in a much shorter time and at a much lower cost than individual experts who have spent years building up their knowledge.

This will rapidly lead to increased competition and a changed outlook in the markets. Established business models will need to undergo an at times painful transformation process. In addition, the skills requirements in the labour market will alter, i.e. due to the increased growth caused by the digital revolution, employers in business, academia and politics will demand better qualifications from prospective employees. With growing demand for big data methods there will be lucrative opportunities for career changers with a background in statistics, mathematics, information technology, data analysis, artificial intelligence or robotics, as they have sought-after transferable skills.

Many unanswered questions

For all the promise and positive effects associated with big data, some central questions remain unanswered. What influence are the new technologies and analysis methods going to have on our everyday lives? Many sectors are already going through painful adjustment processes. Which other industries will be affected, and which digitisation strategies could they use to counteract the competitive disadvantages they face? How will the stakeholders and the citizens deal with the imbalance in data sovereignty? What role will Germany play in an increasingly digital, globally networked world? With its knowledge-intensive economy and lack of natural resources, a country like Germany is ideally placed to play a leading international role in cutting-edge internet technologies, especially in the area of IT security.

The digital world needs rules

The commercial and technological foundations for the debate about big data are already in place. Now the time has come to discuss essential data protection aspects based on fundamental rights and freedoms, as well as the potential threat of competition distortion from the digital ecosystems. The latter aspect is particularly important, because the corrective market consolidation in the digital world of the large internet corporations is happening much faster than in, for example, the automotive sector. Decision-makers from business, politics and academia, as well as civil rights groups, must now play an active part in shaping the necessary legislative framework. This goes particularly for the question of how to achieve a balance between data protection legislation on purpose limitation and the big data methods currently being practised.

The many different interests must be discussed now

It is only a matter of time before we reach the 'Internet of Things'; sooner or later humanoid robots will be providing increased efficiency and support in many households. In future, people are going to be interacting with their computers (artificial intelligence) in order to optimise their day-to-day lives. The burning question in this regard is not when this is going to happen, but rather how future generations will handle modern technologies, what general parameters they are going to agree on, what boundaries they will set, and whether people will be able to trust that their rights are going to be enforceable.

The wide spectrum of different hopes and much-vaunted potential – but also the well-justified concerns – show just how multi-faceted the developments and

applications of big data are. After all is said and done, big data can be used for a vast range of different commercial, reputable purposes, but equally to criminal or at least highly dubious ends. For every positive example there is also a worrying but equally realistic scenario.

Danger of losing trust

A favourable legislative environment, ideally with conditions that apply on an international basis, would aim to create the necessary equilibrium between stimulating the development opportunities for big data without putting too much emphasis on the risks. This balancing act will not be an easy one, but it ought to be tackled at this early stage of development, and at an international level. There are not going to be any perfect solutions or definitive answers for many of the future problems and questions relating to big data. The challenge is to find a way of integrating modern technologies and methods into people's everyday lives in a useful way without

— violating any civil rights and liberties or democratic principles,

— without discrimination or manipulation,

— and without making people more afraid of using digital channels because different stakeholders are conducting mass-surveillance of private data.

It is up to us, as the valuable, creative suppliers of the ideas behind every (r)evolutionary technology or innovation, to keep a tight hold of the reins and set an appropriate course. May the Force be with us.

Thomas F. Dapp (+49 69 910 31752, thomas-frank.dapp@db.com)

Veronika Heine

Reference list

Bahr, F. et al. (2012). Schönes neues Internet? Chancen und Risiken für Innovation in digitalen Ökosystemen. (Great new internet? Opportunities and risks for innovation in digital ecosystems.) Policy Brief. Stiftung Neue Verantwortung. Berlin.

Beck, S., Grzegorzek, M. et al. (2013). Mit Robotern gegen den Pflegenotstand. (Using robots to alleviate the care crisis.) Policy Brief 04/13. Stiftung Neue Verantwortung. Berlin.

Big Data – Vorsprung durch Wissen. Innovationspotenzialanalyse. (Big Data – Getting ahead through knowledge. Analysis of innovation potential.) 2012. Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS). Sankt Augustin near Bonn.

Boyd, D., Crawford, K. (2013). Critical questions for Big Data. Provocations for a cultural, technological and cultural phenomenon. In Geiselberger, H., Moorstedt, T. (2013). Big Data. Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.) eBook Suhrkamp Verlag Berlin.

Brown, B. et al. (2011). Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Dapp, T. (2012). Homo biometricus: Biometric recognition systems and mobile internet services. Current Issues. Deutsche Bank Research. Frankfurt am Main.

Dapp, T. (2013). The future of (mobile) payments: New (online) players competing with banks. Current Issues. Deutsche Bank Research. Frankfurt am Main.

Dapp, T. and Heymann, E. (2013). Dienstleistungen 2013. Heterogener Sektor verzeichnet nur geringe Dynamik. (Services 2013. Heterogeneous sectors record only slow growth.) Aktuelle Themen. Deutsche Bank Research. Frankfurt am Main.

Geiselberger, H., Moorstedt, T., (2013). Big Data: Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.) eBook Suhrkamp Verlag Berlin.

Ginsberg, Jeremy et al. (2009). Detecting influenza epidemics using search engine query data. Google Inc. Centers for Disease Control and Prevention in *Nature* Vol. 457.

Hayek, F. A. von (2004). Der Weg zur Knechtschaft. (The Road to Serfdom.) 4th edition. Mohr Siebeck Verlag. Tübingen.

Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. (Small data, big impact. Big data in a nutshell.) DigitalKompakt LfM #06. Landesanstalt für Medien Nordrhein-Westfalen (LfM). Düsseldorf.

Kranzberg, M. (1986). Technology and history: Kranzberg's laws. In Technology and Culture. 27/2. P. 544-560.

Manig, M. and Giere, J. (2012). Quo Vadis Big Data – Herausforderungen – Erfahrungen – Lösungsansätze. (What next for big data – challenges – experiences – approaches.) TNS Infratest GmbH. Munich.

Mayer-Schönberger, V., Cukier, K. (2013). Big Data. Die Revolution, die unser Leben verändern wird. (Big Data. The revolution that will change our life.) Redline Verlag. Munich.

Morozov, E. (2013). Big Data. Warum man das Silicon Valley hassen darf. (Why you're allowed to hate Silicon Valley.) Frankfurter Allgemeine Zeitung review supplement of 10 November 2013.

Schaar, P. (2013). Big Brother und Big Data – Was heißt eigentlich Datenschutz auf Amerikanisch? (Big Brother and Big Data – How do you say 'data protection' in American?) Talk given as part of the 41st Römerberg debates on 26 October 2013 in Frankfurt am Main.

Schroeck, M. et al. (2012). Analytics: Big Data in der Praxis. Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. (Analytics: practical uses of big data. How innovative companies are drawing benefit from their data.) IBM Institute for Business Value in partnership with the University of Oxford's Saïd Business School.

Weichert, T. (2013). Big Data – eine Herausforderung für den Datenschutz. (Big Data – a challenge for data protection.) In Geiselberger, H., Moorstedt, T. (2013). Big Data. Das neue Versprechen der Allwissenheit. (Big Data: The new promise of omniscience.) eBook Suhrkamp Verlag Berlin.

Weichert, T. (2013). Big Data und Datenschutz. (Big data and data privacy.) Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein. (Independent Regional Centre for Data Privacy, Schleswig-Holstein).

Further reading

Becker, K., Stalder, F. (2010). Deep Search. Politik des Suchens jenseits von Google. (Deep Search. The politics of search beyond Google) Federal Agency for Civic Education. www.bpb.de. Bonn.

Bloching, B. et al. (2012). Data Unser. Wie Kundendaten die Wirtschaft revolutionieren (The Store's Prayer. How customer data is revolutionising the economy.) Redline Verlag. Munich.

Kurz, C., Rieger, F. (2012). Die Datenfresser. Wie Internetfirmen und Staat sich unsere persönlichen Daten einverleiben und wie wir die Kontrolle darüber zurückerlangen. (The data devourers. How internet firms and the state are feasting on our personal data and how we can regain control of this.) Fischer Verlag GmbH. Frankfurt am Main.

Schmidt, J., Weichert, T. (2012). Datenschutz. Grundlagen, Entwicklungen und Kontroversen. (Data privacy. Fundamentals, developments, controversies.) Federal Agency for Civic Education. Bonn.

Shapiro, C. and Varian, H. (1999). Information Rules. A Strategic Guide to the Network Economy. Havard Business Review Press, Boston Massachusetts.

Deutsche Bank
Research

Focus topic Germany

Our publications can be accessed, free of charge, on our website www.dbresearch.com You can also register there to receive our publications regularly by E-mail.

Ordering address for the print version:
Deutsche Bank Research
Marketing
60262 Frankfurt am Main
Fax: +49 69 910-31877
E-mail: marketing.dbr@db.com

Available faster by E-mail:
marketing.dbr@db.com