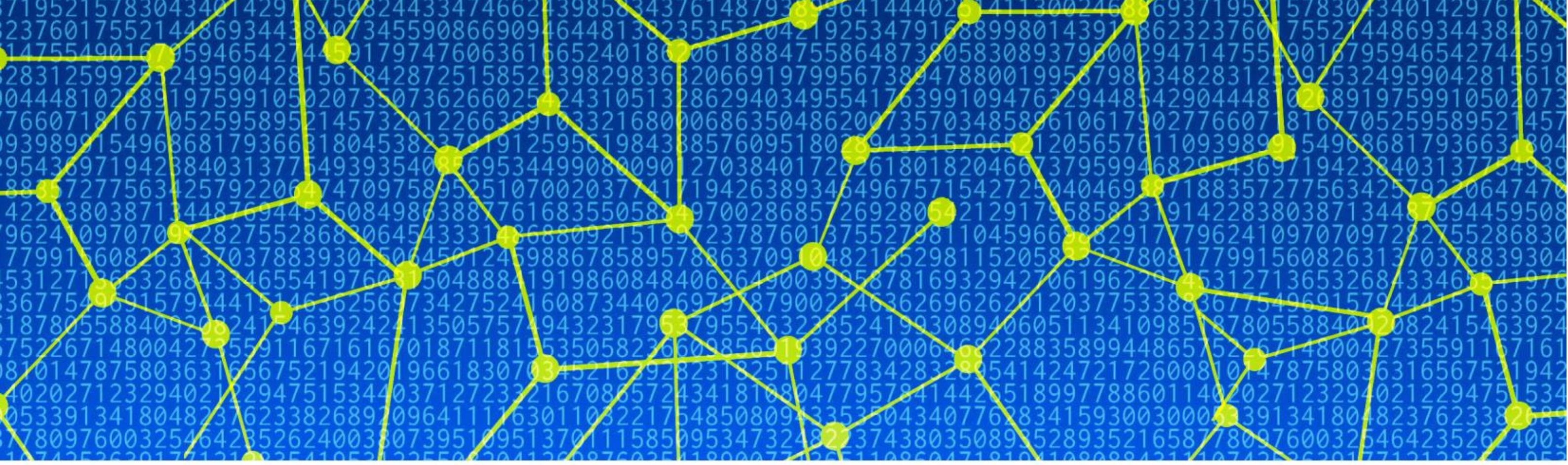


European Data Protection Supervisor

Opinion 7/2015

Meeting the challenges of big data

A call for transparency, user control, data protection by design and accountability

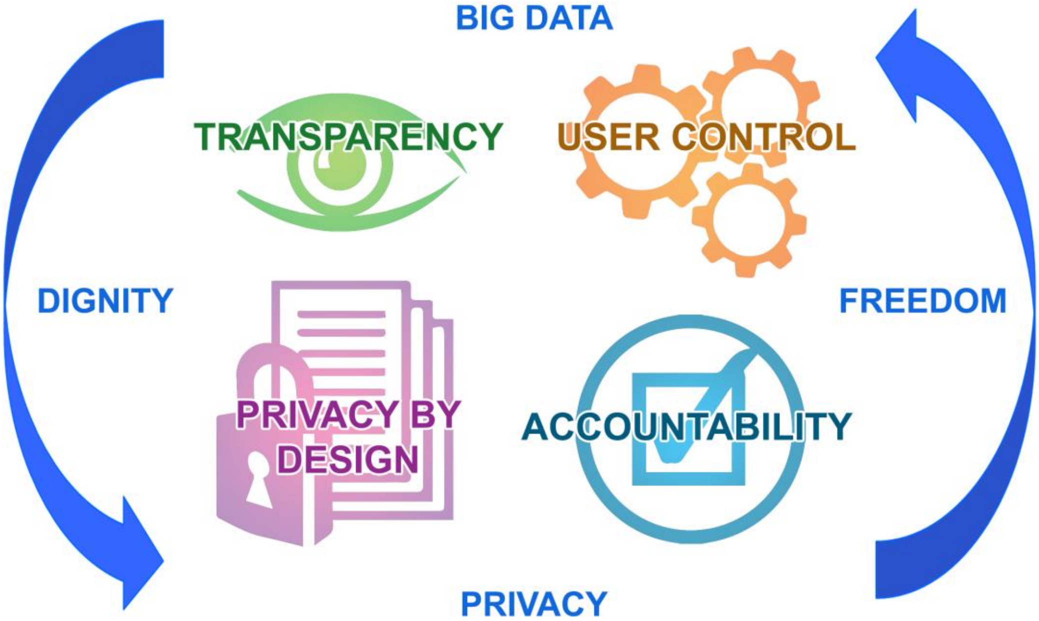
EDPS

19 November 2015

The European Data Protection Supervisor (EDPS) is an independent institution of the EU. The Supervisor is responsible under Article 41.2 of Regulation 45/2001 'With respect to the processing of personal data… for ensuring that the fundamental rights and freedoms of natural persons, and in particular their right to privacy, are respected by the Community institutions and bodies", and "…for advising Community institutions and bodies and data subjects on all matters concerning the processing of personal data'.

The Supervisor and Assistant Supervisor were appointed in December 2014 with the specific remit of being more constructive and proactive, and they published in March 2015 a five-year strategy setting out how they intended to implement this remit, and to be accountable for doing so.

This Opinion follows on from the EDPS's previous Opinion on Digital Ethics[1]*. It addresses the challenge of data protection to 'go digital' -the first objective of the EDPS Strategy– 'customising existing data protection principles to fit the global digital arena', also in the light of the EU's plans for the Digital Single Market. It is consistent with the approach of the Article 29 Working Party on data protection aspects of the use of new technologies, such as the 'Internet of Things', to which the EDPS contributed as a full member of the group.*


BIG DATA
TRANSPARENCY
USER CONTROL
DIGNITY
FREEDOM
PRIVACY BY DESIGN
ACCOUNTABILITY
PRIVACY

‘The right to be let alone is indeed the beginning of all freedom’[2].

Big data, if done responsibly, can deliver significant benefits and efficiencies for society and individuals not only in health, scientific research, the environment and other specific areas. But there are serious concerns with the actual and potential impact of processing of huge amounts of data on the rights and freedoms of individuals, including their right to privacy. **The challenges and risks of big data therefore call for more effective data protection.**

Technology should not dictate our values and rights, but neither should promoting innovation and preserving fundamental rights be perceived as incompatible. New business models exploiting new capabilities for the massive collection, instantaneous transmission, combination and reuse of personal information for unforeseen purposes have placed the principles of data protection under new strains, which calls for thorough consideration on how they are applied.

European data protection law has been developed to protect our fundamental rights and values, including our right to privacy. **The question is not *whether* to apply data protection law to big data, but rather *how* to apply it innovatively in new environments.** Our current data protection principles, including transparency, proportionality and purpose limitation, provide the base line we will need to protect more dynamically our fundamental rights in the world of big data. They must, however, be complemented by ‘new’ principles which have developed over the years such as accountability and privacy by design and by default. The EU data protection reform package is expected to strengthen and modernise the regulatory framework[3].

The EU intends to maximise growth and competitiveness by exploiting big data. But the Digital Single Market cannot uncritically import the data-driven technologies and business models which have become economic mainstream in other areas of the world**.** Instead it needs to show leadership in developing accountable personal data processing. The internet has evolved in a way that surveillance - tracking people’s behaviour - is considered as the indispensable revenue model for some of the most successful companies. This development calls for critical assessment and search for other options.

In any event, and irrespective of the business models chosen, organisations that process large volumes of personal information must comply with applicable data protection law. The European Data Protection Supervisor (EDPS) believes that responsible and sustainable development of big data must rely on **four essential elements**:

- organisations must be much more **transparent** about how they process personal data;
- afford users a higher degree of **control** over how their data is used;
- **design** user friendly data protection into their products and services; and
- become more **accountable** for what they do.

When it comes to **transparency**, individuals must be given clear information on what data is processed, including data observed or inferred about them; better informed on how and for what purposes their information is used, including the logic used in algorithms to determine assumptions and predictions about them.

User control will help ensure that individuals are more empowered to detect better unfair biases, to challenge mistakes. It will help prevent the secondary use of data for purposes that do not meet their legitimate expectations: With a new generation of user control, individuals will, where relevant, be given more genuine and better informed choice and enjoy greater possibilities themselves to use their personal data better.

Powerful **rights of access and to data portability and effective opt-out mechanisms** may serve as a **precondition to allow users more control over their data**, and may also help contribute to the development of new business models and more efficient and transparent use of personal data.

By **building data protection into the design of their systems and processes**, and **adjusting** data protection to **allow more genuine transparency and user control, accountable controllers** will also be able to benefit from the advantages of big data while at the same time ensuring that individuals' dignity and freedoms are respected.

But data protection is only part of the answer. The EU needs to deploy in a more coherent way the modern tools available, including in the area of **consumer protection, antitrust, research and development**, to ensure safeguards and choice in the marketplace where privacy friendly services can thrive.

In order to answer the challenges of big data we **need to allow innovation and protect fundamental rights at the same time.** It is now up to companies and other organisations that invest a lot of effort into finding innovative ways to make use of personal data to use the same innovative mind-set when implementing data protection law.

Building on previous contributions by academia and many regulators and stakeholders, **the EDPS wants to stimulate a new open and informed discussion in and outside the EU**, by better involving civil society, designers, companies, academics, public authorities and regulators on how best to use industry's creative potential to implement the law and safeguard our privacy and other fundamental rights the in best possible way.

TABLE OF CONTENTS

1. Big data analytics: opportunities, risks and challenges

1.1 ‘Big data’ and ‘big data analytics’

In general terms, as a common denominator of the various definitions available, ‘big data’[4] refers to the practice of combining huge volumes of diversely sourced information and analysing them, using more sophisticated algorithms to inform decisions. Big data relies not only on the increasing ability of technology to support the collection and storage of large amounts of data, but also on its ability to analyse, understand and take advantage of the full value of data (in particular using analytics applications).

The expectation from big data is that it may ultimately lead to better and more informed decisions. For instance, big data may bring better insights in scientific and medical research, increased self-knowledge for individuals, products, services and medical treatments that are more personalised and thus better suited to the individual, and better automated decisions for businesses and other organisations that process data. These automated decisions, in turn, may lead to increased efficiency, with promises of various commercial and other applications.

The information processed by big data applications is not always personal: data generated by sensors for monitoring natural or atmospheric phenomena like the weather or pollution, or for monitoring technical aspects of manufacturing processes, may not relate to ‘*an identified or identifiable natural person*’. But one of the greatest values of big data for businesses and governments is derived from the monitoring of *human* behaviour, collectively and individually, and resides in its predictive potential.

One result is the emergence of a revenue model for Internet companies relying on tracking online activity. Such ‘big data’ should be considered personal even where anonymisation techniques have been applied: it is becoming and will be ever easier to infer a person’s identity by combining allegedly ‘anonymous’ data with publicly available information such as on social media. Furthermore, with the advent of the ‘Internet of Things’, much of the data collected and communicated by the increasing number of personal and other devices and sensors will be personal data: the data collected by them can be easily related to the users of these devices whose behaviour they will monitor. These may include highly sensitive data including health information and information relating to our thinking patterns and psychological make-up.

Big data applications that process personal data often evaluate some aspects of individuals, including health or financial risks. In other cases, businesses use big data in order to market products or services to us in a more efficient and effective way and/or to provide a more personalised service. An increasing number of other applications also rate individuals for various purposes: an employer may tap into big data to pre-select the most promising candidates for a vacancy, and travellers use apps to see which taxi companies or bed and breakfast hosts provide the best service. In yet other cases, organisations need our data for research of one sort or another: they would like to detect general trends and correlations in the data[5].

1.2 What are today the main risks and challenges of big data?

The application of big data offers significant benefits for individuals and society but also raises serious concerns about its potential impact on the dignity and the rights and freedoms of individuals, including their right to privacy. These risks and challenges have already been extensively analysed by data protection experts worldwide[6], therefore the EDPS only highlights a few of the key concerns.

Lack of transparency. While the complexity of data processing increases, organisations often claim secrecy over 'how' data is processed on grounds of commercial confidentiality. As indicated in a 2014 report by the US White House, '*some of the most profound challenges revealed during this review concern how big data analytics may ... create such an opaque decision-making environment that individual autonomy is lost in an impenetrable set of algorithms'*[7]. Unless individuals are provided with appropriate information and control, they '*will be subject to decisions that they do not understand and have no control over*'[8]. Individuals cannot efficiently exercise control over their data and provide meaningful consent in cases where such consent is required. This is all the more so as the precise future purposes of any secondary use of the data may not be known when data is obtained: in this situation, controllers may be unable or reluctant to tell individuals what is likely to happen to their data and to obtain their consent when required.

Informational imbalance between the organisations who hold the data and the individuals whose data they process is likely to increase with the deployment of big data applications[9].

Failing to address these issues may create the risk of core principles of data protection being compromised. The perceived opportunities in big data provide incentives to collect as much data as possible and to retain this data as long as possible for yet unidentified future purposes. Some advocates of big data demand derogations from the central principles, particularly those of purpose limitation and data minimisation, and argue that these principles should not (or should not be fully) applied to big data processing. Big data also challenges the principles of accuracy and relevance of data. Big data applications typically tend to collect data from diverse sources, and without careful verification of the relevance or accuracy of the data thus collected.

One of the potentially most powerful uses of big data is to make predictions about what is likely to happen but has not yet happened and what we are likely to do but have not yet done. For example, big data might be used to predict a child's performance at school or an adult's susceptibility to illness or premature death, to default on credit or commit crime. Notwithstanding the potential benefits, data has been described by one commentator as the 'pollution problem of the information age'[10], with the a risk of a **'dictatorship of data'** where, according to one study by a European data protection authority, *'we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicate our probable actions may be'*[11].

The expected benefits of statistics based prediction may further increase overconfidence in its capabilities. **Big data applications may find spurious correlations in data**, even in cases where there is no direct cause and effect between two phenomena that show a close correlation. In these cases there is a risk of drawing inaccurate but also – when applied at the individual level – **potentially unfair and discriminatory conclusions**.

These and other characteristics of big data, extensive use of automated decisions and predictive analysis may also lead to broader undesirable changes in the development of our societies. Importantly, they may lead to discrimination, re-enforcement of existing stereotypes, **social and cultural segregation and exclusion**[12].

The accumulation of massive personal data sets which feed big data analytics is possible because of the constant, invisible tracking of online activity. This surveillance may also have a **chilling effect on creativity and innovation**.

Big data analytics are used to identify behaviour which statistically speaking poses less risk to generates more value for organisations processes the data. There is a tendency to discourage or penalise spontaneity, experimentation or deviation from the statistical ‘norm’, and to reward conformist behaviour. For example, the banking and insurance sectors have an obvious interest in acquiring granular insights into the risk posed by an individual which might be revealed by combinations of datasets generated by activity on social media and by connected devices tracking location and other personal information data and the increasing number of connected objects. The need for a loan or insurance could nudge or coerce individuals into avoiding contact with certain people or companies or visiting areas with high crime rates the same way as it makes people installing ‘black boxes’ which allows external controller to monitor them while they are driving[13].

The very fact that our behaviour is constantly tracked and analysed may also caution us to watch how we behave and encourages us to conform, in advance, to what we perceive to be as the expected norm. These trends can also have a chilling effect on freedom of expression and other activities necessary to maintain a democratic society such as exercising the rights of free assembly or association.

Looked at this way, the rights to privacy and to the protection of personal data are a precondition for individuals to develop their personalities and to lead their own lives as independent human beings, as well as a precondition to exercise cherished rights and freedoms, and indeed, also a **precondition for individuals and also for society to innovate**.

In order to ensure that our fundamental rights and values, including our ability for us as a society and as individuals to continue to innovate, are safeguarded and maintained, big data must be deployed in a more responsible and sustainable manner. As argued in our previous Opinion 4/2015, there is an urgent need to address a Big Data Protection Ecosystem consisting of:

- organisations which are much more transparent about how they process personal data;
- individuals that are able to benefit from a higher degree of control over how their data is used;
- data protection designed into products and services; and
- more accountable controllers.

Each of these four themes will be briefly addressed in this Opinion.

2. Transparency: end covert profiling

2.1. Disclosing the logic involved in Big Data analytics

Transparency of automated decisions is taking an increasingly important role with the advent of big data analytics. Disclosing the logic of decision-making can help individuals better to verify whether the conclusions drawn by the organisations processing the data and impacting the individuals are accurate and fair. They can better understand, and perhaps rectify, the criteria underpinning, and the factors influencing the decision.

As a society, we must be able to look into the 'black box' of big data analytics in order to ensure that any particular analytics application can be safely deployed and will benefit us all[14]. Accordingly, organisations are expected to disclose the logic involved in big data analytics where there is an effect (direct or indirect) on the individual. They must do so proactively[15], without individuals having to actively take steps to seek disclosure[16].

The personal data processed in big data contexts are no longer primarily comprised of information that individuals knowingly gave to organisations. Much of the personal data processed is now observed or inferred: recording of online activities and locations of smartphones and tablets and increasing possibilities to track activities in the 'real world' by smart devices and the 'Internet of Things' add to the huge pile of data from which inferences and predictions are made about us. Transparency is also important where data was collected from publicly available sources.

Whether the data are volunteered, observed, or inferred[17], or collected from public sources, individuals are fully entitled to know what they are and from where and from whom the controllers obtained it. It is becoming increasingly necessary to give to the individuals more proactively the data itself, '*in an intelligible form'* as well as the source of the data.

Protecting business confidentiality or trade secrets cannot generally overrule the fundamental rights of individuals to privacy and data protection. Instead, reconciling the two requires a careful balancing[18]. Neither is the decision about disclosure a binary one. Rather, the assessment needs to consider which information can be disclosed and also disclosure and assessment procedures. For example, in some cases trusted third parties as assessors can be used instead of disclosing all details to the individual or the public [19].

Data protection authorities (and other regulators, such as, for example consumer protection authorities, competition authorities, financial and insurance regulators) should also be able to look into the 'black box'.

For these reasons, and in order to ensure an end to covert profiling, we recommend that the provisions of the proposed EU Data Protection Regulation on transparency be reinforced and should specifically include disclosure of the 'logic of decision-making', the data itself, as well as its source.

2.2. Better tools for informing individuals

It is also important to make progress on how to disclose information to the individuals. Any information relating to the processing of personal information must use clear and plain language, tailored to the relevant audience, allowing individuals to make sense of complex information, and be easily accessible. If the processing becomes more complex, data controllers have the responsibility to ensure users and consumers are better informed.

These policies should genuinely serve to safeguard the interests of the individual concerned by personal data processing, not merely to shield the controller from legal liability. As with consumer law, where there is any ambiguity in these policies they should be interpreted in favour of the individual not the controller. They should also be truthful and honest.

More traditional privacy policies, while important for accountability, should not be the only, or even the main source of information for individuals. Data protection authorities have long been recommending a 'layered' notice[20] informing the data subjects about their data being processed step by step. This means providing the individual with the essential information about the processing at the point where the individual needs to make a decision based on the information (for example, an individual needs to know whether an app downloaded will have access to his location data before he chooses to install it), and providing further information in other formats, for example, via more detailed information on a website.

3. Beyond unreadable privacy policies: user control and sharing the benefits of big data with the individuals

At global level, the current debate is polluted by misunderstandings about the concept of notice and consent. Consent under European data protection law[21] has never meant long and impenetrable privacy policies, written by lawyers for lawyers, which users must 'consent' to unless they wish to abandon the use of the desired service altogether. Instead, it means a genuine, freely-given choice with the alternative, without any detriment, to say 'yes'. It also requires a clear understanding of what one agrees to.

Consent is not always required for organisations to process data[22]. However, when consent is required, it should be a genuine one: mere ticking of a box without understanding of what we agree to, and without meaningful choice whether we do so, is not sufficient to signify our consent for complex big data applications. Transparency and user control must become reality[23].

3.1 Short of consent: right to object and opt-out mechanisms

The right to object to processing (which is not frequently exercised in today's practice) can become a powerful tool in the hand of the individuals when it is implemented as an unconditional, 'no questions-asked' opt-out. This may, in some circumstances, help establish the right balance between the right of the individual to have a degree of control over his or her data and the flexibility required for businesses to develop and innovate and make best use of the vast amount of data generated on-line and off-line[24].

An unconditional opt-out means that an individual is aware that his data is processed and knows he could opt-out if he chose to do so. He may or may not wholeheartedly embrace the fact that his data is being processed, however, often is not sufficiently negatively affected -or simply not 'bothered'- to change the default setting. Opt-out subtly influences the individual to agree, without altogether denying him the right to disagree.

Especially in borderline cases where the balance between the legitimate interests of the controller and the rights and interests of the data subjects are difficult to strike, a well-designed and workable mechanism for opt-out, while not necessarily providing data subjects with all the elements that would satisfy a valid consent under European data protection law[25], could play an important role in safeguarding the rights and interests of the individuals[26].

As a society, we must make wise choices to the conditions under which we require controllers to obtain genuine consent and when to content ourselves merely with an assessment of the balance of interests and an opt-out. We must in particular, aim to distinguish data processing whose benefits are general/societal, from those that merely provide economic benefits to those processing the data. We must also assess the potential impact on the individuals concerned, and carefully balance these two as well as all other relevant factors[27].

Opt-out may be facilitated by industry-wide arrangements so long as these will be effective and easy to exercise. However, more efforts are needed before any particular initiatives may be endorsed as experience thus far with such arrangements have led to little in the way of concrete results[28].

3.2 Beyond consent: user control and sharing the benefits

Right of access and data portability

The right to access and correct one’s personal data is one of the fundamental principles of European data protection law[29] and is becoming increasingly more important with advances of big data analytics. Individuals must be empowered to better detect unfair biases and challenge mistakes arising from the logic used in algorithms to determine assumptions and predictions and a strong right of access and correction is a precondition to this.

Nevertheless, only few individuals exercise these rights in practice[30]. One of the reasons why these potentially powerful access rights have not emerged as more powerful tools in practice is because individuals often do not have the time or interest to *‘indulge in transparency and access for their own sake’*[31]. If, however, individuals were given the ability to use their personal data to benefit from it in a tangible way, the situation might change. This could be achieved through ‘featurization’ of data protection: instead of an administrative burden, providing access rights may become a feature of the service provided to the customers[32]. An everyday example is access to one's banking information online.

With big data on the increase, organisations are using personal data for secondary purposes not strictly necessary for the delivery of the services in the first place. If they wish to do so they should also be prepared to share the wealth created by the processing of personal data with those individuals whose data they process[33]. This is a basic requirement of fairness - however, it is not merely an ethical imperative.

Data is often compared with other resources, such as oil, which are traded, ideally by equally well informed parties to the transaction. Markets for personal information, however, are far from being transparent, fair or efficient. Customers are generally unaware of the precise value of the personal data that they give away in exchange of ‘free services’. As a result, they are not fairly compensated for their personal information.

In what way and to what extent individuals should benefit from the wealth created by the processing of their personal data is a key question to reflect upon in the context of the development of the Digital Single Market.

One of the ways to give more control to individuals, share with them the benefits of big data, and at the same time, create incentives for efficient and transparent processing of personal data, is via data portability. Data portability would require that organisations:

- provide individuals with access to their own data in portable, interoperable and machine-readable (in other words, usable and reusable) format,
- allow them to modify, delete, transfer, or otherwise further process their own data,
- allow them to switch providers (e.g. transfer their photos, banking or fitness records, or emails to a different service provider), and
- allow them to take advantage of other third party applications to analyse their own data and draw useful conclusions (*e.g.*, change dietary or exercise habits, get personalized health care, make wiser investment decisions, switch to a cheaper electricity provider).

Allowing data portability could enable businesses and individuals to maximise the benefits of big data in a more balanced and transparent way and may help redress the economic imbalance between controllers on one hand and individuals on the other. It could also let individuals benefit from the value created by the use of their personal data: it could allow them to use the data for their own purposes, or to license the data for further use to third parties, in exchange of additional services, or for cash value. Further, it could also help minimise unfair or discriminatory practices and reduce the risks of using inaccurate data for decision-making purposes.

In addition, data portability is not only good for data protection, but also for competition and consumer protection: In particular, it can foster a more competitive market environment, by allowing customers more easily to switch providers (e.g. in the context of online banking or in case of energy suppliers in a smart grid environment). Further, it can also contribute to the development of additional value-added services by third parties who may be able to access the customers' data at the request and based on the consent of the customers. This, again, may bring down barriers to entry to new markets that require access to personal data, and help create more competitive, less monopolistic market structures[34].

In view of its benefits, the EDPS firmly supports the inclusion of a strong right to data portability in the proposed EU Data Protection Regulation, as well as the inclusion of this right in relevant sectorial legislation, regulation or guidance where appropriate (e.g. with regard to smart metering). We also encourage government or private initiatives to facilitate data portability.

Personal data spaces

Complementing and building on data portability, one method for giving individuals better control over their data, who can access it and for what purpose, could be the use of personal data spaces (also called 'data stores' or 'data vaults'). Examples of continuously updated real-time 'big data' that may be stored in a personal data space may include an individual's location tracked by sensors in his car or mobile phone or blood pressure and other health/fitness-related data tracked by a fitness tracker or a medical device.

The European Commission Communication on Big data[35] specifically refers to and encourages the use of 'personal data spaces'[36] as user-centric, safe and secure places to store and possibly trade personal data. We share the view that innovative digital tools and business models that are based on consumer empowerment should be encouraged. These include mechanisms for individuals to participate in the use and distribution of their information, and which allow them to benefit from such data-sharing.

This could potentially enable a shift from business models where organisations increasingly track individuals' behaviour online and offline without their full knowledge and consent, to one where individuals manage their own information for their own purposes, and share some of this information when they want, with whom they want, and at a fair value and subject to adequate safeguards[37]. Personal data stores could help address some of the concerns over the loss of individual control over personal data that has been highlighted above as one of the key concerns about big data[38].

The EU should examine how to promote reliable, trustworthy, user-friendly and interoperable tools and products and the benefits and constraints and technological challenges[39].

3.3 New, innovative ways to provide information, access and control to individuals

Companies and other organisations that invest a lot of effort into finding innovative ways to make use of personal data should use the same innovative mind-set when designing **new, innovative ways to provide information, access and control to individuals**[40].

New user-friendly ways should be developed and offered to allow individuals to provide or refuse informed consent. Giving individuals some degree of control over data use is often a legal requirement or good practice that will also help controllers create trust.

For example, individuals should be able to effortlessly switch on and off the tracking or information sharing of the devices and applications they use, based on location, time and date, by application as well as globally. Better ways should also be offered to facilitate correcting, updating or deleting data, or modifying who may have access to it, monitoring who actually accessed it, and for what purposes. This brings us to our next topic, data protection and privacy by design.

4. Data protection and privacy by design

Privacy and data protection by design aims at building privacy and data protection into the design specifications and architecture of information and communication systems and technologies. It is not limited to technical aspects, organisational measures are just as important.

Technology and privacy-friendly engineering can play a key role in ensuring that transparency and user control, as outlined above, will become a reality. Laws, regulations, contractual terms, internal procedures, and privacy policies, while important, will not suffice on their own. Individuals need to be offered **new, innovative ways to be informed about what happens to their data, and to exercise control over their data**. This requires innovative and privacy-friendly engineering as well as privacy-friendly organisational arrangements and business practices. Innovative and responsible engineering can facilitate, among others, the exercise of individuals' rights of access, objection, opt-out, correction, as well as data portability. Privacy-friendly engineering can also be invaluable in helping develop new business models for generating value from for example, data stores.

Another, altogether different, area, where innovative engineering solutions are to be encouraged, is related to the concept of **'functional separation'**. In case an organisation processing data only wants to detect trends and correlations in the information rather than directly applying any insights they gained to the individuals concerned, 'functional separation' may potentially play a role in reducing the impact on the rights of individuals, while at the same time allowing organisations to take advantage of secondary uses of data[41]. The objective of functional separation is to take technical and organisational measures to ensure that data used for research purposes cannot then be used to 'support measures or decisions' with regard to the individuals concerned (unless specifically authorised by these individuals)[42].

Further, and despite their limitations, appropriate **anonymisation techniques** can still play a role in ensuring the safe use or sharing of data within an organisation, among different organisations or when data is made publicly available, such as in case of 'open data' projects[43]. Anonymisation of data cannot be achieved by just stripping a dataset of some directly identifying attributes. The bigger and the more comprehensive a collection of data becomes, the more possibilities exist to identify the individuals whom the data relates to, especially when data is retained for longer periods of time and/or shared[44]. Careful use of such techniques, in combination with other safeguards (e.g. restrictions on data retention periods, access control), however, may help ensure compliance with data protection laws in some situations.

Finally, initiatives and investments in the use and deployment of big data must treat appropriate security as a critical pre-condition for socially acceptable use of big data, and must address risk assessment and security measures as an integral part of big data.

5. Accountability

Accountability requires that controllers put in place internal mechanisms and control systems that ensure compliance and provide evidence –including internal policies and audit reports– to demonstrate compliance to external stakeholders, including supervisory authorities. Accountability is not a one-off exercise: regular verification that these internal control systems continue to be fit and any data processing continues to comply with the law is an essential element of accountability.

There are many elements that make up good and accountable practices. Privacy and data protection by design and by default, data protection impact assessments, auditing and certification, and the availability of the right data protection expertise, including a data protection officer, within the organisation may all contribute and form integral part of an accountable internal control system, and should be required and encouraged as appropriate as they may play an important role in ensuring that big data is used responsibly.

It is often a challenging task to decide what is fair and lawful and what is not when it comes to big data analytics.

Some of the key decisions an accountable organisation must make under European data protection law include:

- whether any secondary use of data complies with the principle of purpose limitation,

- whether data initially used in one context can be considered adequate, relevant, and proportionate to be reused in another context, and
- whether, in the absence of obtaining consent from the individuals, an organisation can rely on its legitimate interest to process any data.

While these assessments are based on legal requirements, they often require a comprehensive balancing exercise and consideration of many factors, including whether the data processing meets the reasonable expectations of the individuals concerned, whether it may lead to unfair discrimination or may have any other negative impact on the individuals concerned or on society as a whole. These assessments often raise challenging questions of business ethics and fairness, and cannot be reduced to a simple and mechanical exercise of ticking off compliance boxes. The more powerful computers become, the more acute is the challenge: for example, research has found that computers are more accurate than humans at predicting from 'digital footprints' personality traits, political attitudes and physical health [45].

For these reasons, such assessments may be best tackled by a multidisciplinary group (e.g. computer scientists, engineers, lawyers, data protection officers, statisticians, data scientists, doctors, scientists, marketing, insurance or finance specialists).

'Ethics boards' may, where appropriate, play some part towards more accountable internal procedures. Much like similar bodies in the area of scientific research, they could be providing recommendations or binding decisions within the organisation whether or not particular types of big data analytics may be lawfully and ethically deployed. Other organisational arrangements, however, may be just as effective. What matters is to put in place a compliance framework which will help ensure that the decisions that will ultimately be made about any data processing will be 'ethical', 'fair' and 'lawful'.

6. Next steps: putting the principles into practice

In order to answer the challenges of big data we **need to allow innovation and protect fundamental rights at the same time**. To achieve this, the established principles of European data protection law should be preserved but applied in new ways.

6.1. Future-oriented regulation

Negotiations on the proposed General Data Protection Regulation are in the final stages. We have urged the EU legislators to adopt a data protection reform package that strengthens and to modernise the regulatory framework so that it remains effective in the era of big data by strengthening the individuals' trust and confidence online and in the Digital Single Market [46].

In Opinion 3/2015, accompanied by recommendations for a full text of the proposed Regulation, we made it clear that our current data protection principles, including necessity, proportionality, data minimisation, purpose limitation and transparency must remain key principles. They provide the base line we need to protect our fundamental rights in a world of big data[47].

At the same time, these principles must be strengthened and applied more effectively, and in a more modern, flexible, creative, and innovative way. They must also be complemented by new principles such as accountability and data protection and privacy by design and by default.

Increased transparency, powerful rights of access and data portability, and effective opt-out mechanisms may serve as preconditions to allow users more control over their data, and may also help contribute to more efficient markets for personal data, to the benefit of consumers and businesses alike.

Finally, extending the scope of EU data protection law to organisations targeting individuals in the EU, and equipping data protection authorities with the powers to apply meaningful remedies, including effective fines, as the proposed Regulation would provide, will also be a key requirement to effectively enforce our laws in a global environment. The reform process plays a key role in this respect.

To ensure that the rules are effectively enforced, independent data protection authorities must be equipped not only with legal powers and strong instruments, but also with the resources required to match their capacity with the growth of data driven business.

6.2 How EDPS will advance this debate

Good regulation, while essential, is insufficient. Companies and other organisations that invest a lot of effort into finding innovative ways to make use of personal data should use the same innovative mind-set when implementing data protection principles. Data protection authorities, in turn, should enforce and reward effective compliance, and avoid imposing unnecessary bureaucracy and paperwork.

The EDPS, as announced in the EDPS Strategy 2015-2010, aims to contribute to fostering these efforts.

We intend to establish an external ethics advisory group composed of distinguished and independent personalities with a combined experience in multiple disciplines that can '*explore the relationships between human rights, technology, markets and business models in the 21st century', * analyse the impact of big data in depth, assess the resulting changes of our societies and help indicate the issues that should be subject to a political process[48].

We will also develop a model for honest information policies for EU bodies offering online services which can contribute to best practice for all controllers.

Finally, we will also facilitate discussions, for example, with the view to identify, encourage and promote best practice to increase transparency and user control and explore opportunities or personal data stores and data portability. The EDPS intends to organise a Big Data Protection workshop for policy makers and persons handling large volume of personal information in the EU institutions and external experts and to identify where further specific guidance is needed and to facilitate the work of the Internet Privacy Engineering Network ('IPEN') as interdisciplinary knowledge hub for engineers and privacy experts.

Brussels, 19 November 2015

Giovanni BUTTARELLI

European Data Protection Supervisor

Notes

[1] EDPS Opinion 4/2015.

[2] *Public Utilities Commission v. Pollak*, 343 U.S. 451, 467 (1952) (Justice William O. Douglas, dissenting).

[3] On 25 January 2012, the European Commission adopted a package for reforming the European data protection framework. The package includes (i) a 'Communication' (COM(2012)9 final), (ii) a proposal for a general 'Data Protection Regulation' ('proposed Regulation') (COM(2012)11 final), and (iii) a proposal for a 'Directive' on data protection in the area of criminal law enforcement (COM(2012)10 final).

[4] 'Big data refers to the exponential growth both in the availability and in the automated use of information: it refers to gigantic digital datasets held by corporations, governments and other large organisations, which are then extensively analysed (hence the name: analytics) using computer algorithms'; Article 29 Working Party (WP29) Opinion 3/2013 on purpose limitation.

[5] Below are a few examples of big data applications that use personal data (the examples illustrate what technology is capable of, and not what is necessarily ethical or legal to do):

- Medical research and personalised medicine. If scientists are given access to our genetic profiles, medical histories and lifestyle data (e.g. via mobile health and fitness apps, social network information, loyalty and credit card data), using big data analytics on these vast and valuable datasets could potentially revolutionize medical research by allowing scientists to find new correlations, and ultimately, perhaps find new cures to diseases. Big data analytics may also predict whether a patient is likely to be susceptible to a disease, vulnerable to an adverse reaction, or responsive to certain types of medical treatments. This, in turn, may allow doctors to provide personalised, and thus more effective, medical treatment.
- Search engines are built on big data and so are many other online services engaged in the business of rating or recommending content, products or services. Behavioural and targeted advertisement, customised offers and discounts as well as personalised recommendations of media content, hotels, or restaurants are all based on big data.
- Credit scoring takes advantage of big data to assess the risks of failing to pay our debts.
- Combatting tax fraud: if allowed to access certain data from other government agencies or from private businesses, tax authorities, using big data, can cross-reference tax databases with other information such as vehicle registrations, credit card information, or information held by financial intermediaries, to find individuals whose spending/investment patterns and tax contributions do not match up.
- Fight against terrorism and organised crime: intercepting the communications data of large number of people (whether suspect or not), and sifting through them by powerful analytics, intelligence agencies hope to detect terrorist attacks in the making.

[6] See, for example:

- Resolution on Big Data adopted in October 2014 by the 36th International Conference of Data Protection and Privacy Commissioners ('International Conference resolution on big data');
- International Working Group on Data Protection in Telecommunications, Working Paper on Big Data and Privacy (55th Meeting, 5–6 May 2014, Skopje) ('Berlin Group Working Paper on Big Data');
- Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU, adopted on 16 September 2014 (WP 221) ('WP29 Statement on big data');
- UK Information Commissioner's Office, Big data and data protection guide, July 2014 ('ICO guide on big data')

- Norwegian Data Protection Authority's report 'Big data-privacy principles under pressure' 2013 ('Norwegian big data report').

[7] See Big Data: Seizing Opportunities, Preserving Values, Executive Office of the President , May 2014, page 10.

[8] WP29 Opinion 3/2013 on purpose limitation, Annex 2.

[9] Misuses of an imbalance of power can take various forms. Price discrimination is one of them. It enables companies to offer goods or services at different prices to different people, in an effort to extract the maximum price that each consumer is willing to pay. Large datasets on individual behaviour are now readily available, and contain information potentially useful for person-specific pricing. Targeting vulnerable customers is another common form of misuse.

[10] Bruce Schneier, *Data and Goliath*, 2015, p. 238.

[11] Norwegian big data report, page 7, point 8.

[12] In addition, big data can create 'filter bubbles' (or personal 'echo-chambers') for the individuals. In our increasingly personalised world, algorithms make guesses about what information each of us would like to see based on what is known about us (for example, our location, past click behaviour and search and purchase history). Any information we receive is filtered in increasingly more complex and opaque ways. The danger is that we will be less likely to come across information that challenges our existing viewpoints. We will be increasingly effectively isolated in our own cultural and ideological bubbles and divided from the rest of society.

[13] The US National Consumer Law Center found that credit products sold on the basis of non-traditional data-led processes required annual percentage rates of between 134% and 748%; Big Data: A Big Disappointment for Scoring Consumer Credit Risk, March 2014. In August 2015 a US patent was acquired which included technology for examining the credit rating of members of the individuals social network connected to the individual in order to determine whether to process or reject a loan application; 'Facebook patent: Your friends could help you get a loan – or not'; (04.08.2015) http://money.cnn.com/2015/08/04/technology/facebook-loan-patent/. In October 2015 the European Banking Authority, European Securities and Markets Authority, and European Insurance and Occupational Pensions Authority launched a joint investigation in the risks and benefits of big data.

[14] On the concept of 'black boxes' and the importance of transparency, see, for example, 'The Black Box Society, The Secret Algorithms That Control Money and Information' by Frank Pasquale (Harvard University press, 2015).

[15] Under Articles 10 and 11 of Directive 95/46/EC. See also Article 15.

[16] Some everyday examples where 'the logic of decision-making' should be disclosed include a personalised car insurance scheme (using car sensor data to judge driving habits); credit scoring services; a pricing and marketing system that determines how much discount an individual will receive, or what media content to recommend to an individual.

[17] Inferred data also includes an individual's profile, such as, for example, his credit score or the outcome of an assessment regarding his state of health.

[18] In the assessment it should also be considered that secrecy is not the only way to protect genuinely innovative products and services. Indeed, many genuinely innovative algorithms may also be protected by intellectual property rights, rather than relying on business secrets. Patents, for example, offer a high level protection for intellectual property, while at the same time, ensuring more transparency towards the individuals concerned.

[19] On 'qualified transparency', see, e.g. Frank Pasquale: The Black Box Society, p. 160-165.

[20] See, for example, WP29 Opinion 10/2004 on More Harmonised Information Provisions (WP100), WP29 Opinion 2/2009 on the protection of children's personal data (General Guidelines and the

special case of schools) (WP160), and WP29 Opinion 03/2013 on purpose limitation, page 16, and examples 9-10 and 11 in Annex 3, pages 52 and 53.

[21] See Article 7(a) of Directive 95/46/EC.

[22] WP29 Opinion 6/2014 on legitimate interests provides guidance and a set of criteria to help determine in which cases an organisation may rely on the legitimate interest ground and in which cases it must obtain consent from the individuals concerned.

[23] Further, often protection of privacy is falsely equated with choice over whether or not individuals receive targeted online advertising, and user friendly 'dashboards' offer an illusion of control without actually enabling individuals to opt out of tracking and still enjoy the benefits of the internet. It is not just targeted advertising itself which may result in a violation of privacy; but rather, the inability of individuals to avoid being tracked in the first place.

[24] If an organisation that wishes to rely on Article 7(f) of Directive 95/46/EC (legitimate interest) as a legal basis of the processing of personal data must provide individuals with the right to object under Article 14(a) of the Directive under certain conditions. In addition to the right to object under Article 14(a), however, an organisation may also decide to offer a broader, generally applicable, unconditional right to opt-out to individuals. In case of direct marketing, such an opt-out is already a mandatory legal requirement under Article 14(b) of Directive 95/46/EC. This is the type of 'no-questions-asked' opt-out that we are talking about in this Section. See also Opinion 6/2014 of the WP29 on legitimate interests, pages 44 and onwards under the heading *'The right to object and beyond'*.

[25] Under Article 7(a) of Directive 95/46/EC.

[26] WP29 Opinion 6/2014 on legitimate interests provides guidance and examples but emphasises the importance of a case by case assessment. See especially pages 31-33, pages 44-47 as well as examples 4 and 5 in Annex 2 of the Opinion, page 59.

[27] Idem.

[28] There has been much criticism, in particular, of industry initiatives for opt-out from on-line behavioural advertisement and do not track. See, e.g. http://www.nytimes.com/2014/12/27/opinion/the-slow-death-of-do-not-track.html?_r=0.

[29] See Article 8(2) of the Charter of Fundamental Rights of the European Union and Article 12 of Directive 95/46/EC.

[30] See e.g. p. 26 onwards, Omer Tene and Jules Polonetsky (2012), 'Big Data for All: Privacy and User Control in the Age of Analytics, 11 Northwestern Journal of Technology and Intellectual Property 239 (2013), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149364.

[31] Idem.

[32] Idem.

[33] This is in addition to, and beyond what should be a common sense imperative of minimising and compensating for any negative externalities that big data businesses may create - such as, for example, increased security risks for the individuals whose data they process.

[34] See also para 26 of the EDPS Preliminary Opinion on Privacy and Competitiveness in the Age of Big Data, adopted on 26 March 2014. Further, the benefits of data portability are also highlighted by the WP29 in its Opinion 3/2013 on purpose limitation and Opinion 6/2014 on legitimate interest. Both Opinions also specifically refer to initiatives such as 'midata' in the UK (and its equivalent in France), which are based on the key principle that data should be 'released back' to consumers so that they could use it for their own purposes. For more information on midata in the UK and similar initiatives in France, see

http://www.midatalab.org.uk/, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/34747/12-983-midata-company-briefing-pack.pdf and http://mesinfos.fing.org/.

[35] Commission Communication 'Towards a thriving data-driven economy', which sets forth the Commission's strategy on big data COM(2014) 442 final.

[36] Section 4.2.3.1, para 4.

[37] See, for example, Doc Searls, *The Intention Economy: When Customers Take Charge* (Boston: Harvard Business Review Press, 2012).

[38] See, e.g. Ira S. Rubinstein, Big Data: The End of Privacy or a New Beginning? International Data Privacy Law, 2013, Vol 3, No 2.

[39] Issues to be resolved include security, liability and technical feasibility. It would also be necessary to clarify in whose interests personal data spaces act. Finally, it is important to note that personal data spaces should not 'sell' personal data, rather, allow third parties to 'use' personal data, for specific purposes, and specifics periods of time, subject to terms and conditions identified by the individuals themselves, and all other data protection safeguards.

[40] Consent should be sufficiently granular and - even if it is the same entity that further processes the data - should cover the different data processing activities for each service. Combining data for a different purpose also requires a specific consent.

[41] See pages 27, 29, 30 and Annex 2, page 46 of WP29 Opinion 3/2013 on purpose limitation.

[42] There is little evidence of experience with effective implementation of functional separation outside some specialist organisations such as national statistical offices and research institutions. In order to take full advantage of secondary uses of data, it is essential that other organisations develop their expertise and offer comparable guarantees against misuse of data.

[43] For more information how to assess when it is safe to publish aggregated datasets as open data, see Section 6 of WP29 Opinion 06/2013 on open data and public sector information ('PSI') reuse.

[44] An analysis of currently available anonymisation techniques has been provided by the WP29 in its Opinion 5/2014on Anonymisation Techniques.

[45] Computer-based personality judgments are more, accurate than those made by humans, Wu Youyoua, Michal Kosinski, and David Stillwell. December 2014.

[46] EDPS Opinion 3/2015.

[47] We must resist the temptation to water down the current level of protection in an attempt to accommodate a perceived need for a more lax regulatory approach when it comes to big data. Data protection must continue to apply to processing in its entirety, including not only use of the data but also its collection. There is also no justification for blanket exceptions for processing of pseudonymous data or for processing publicly available data. The definition of personal data must remain intact but could do with further clarifications in the text of the Regulation itself. Indeed, it must cover all data that relate to any individual who is identified, singled out, or may be identified or singled out - whether by the data controller or any other party.

[48] EDPS Opinion 4/2015.