Samenvatting Data Analytics & Privacy

Week 1 – Koole – Business Analytics – chapter 1

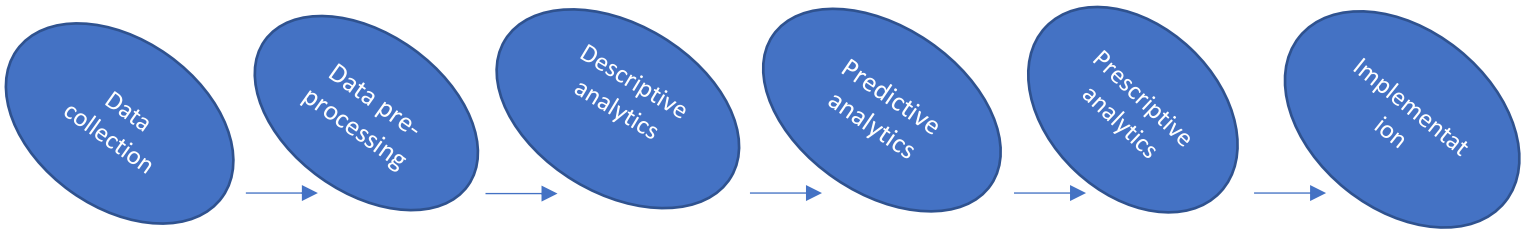
What is business analytics?

Business analytics is a rational, fact-based approach to decision making. These facts come from data, therefore BA is about the science and the skills to turn data into decisions. The science is mostly statistics, AI (data mining and machine learning) and optimization; the skills are computer skills, communication skills, project and change management etc.

Business analytics is often subdivided into three consecutive activities:

1. **Descriptive analytics:** during the descriptive phase, data is analyzed and patterns are found;
2. **Predictive analytics:** the insights from the descriptive phase are consequently used in this phase to predict what is likely to happen in the future, if the situation remains the same
3. **Prescriptive analytics:** in this phase, alternative decisions are determined that change the situation and which will lead to desirable outcomes.

Analytics can only start when there is data. Certain organizations already have a centralized data warehouse in which relevant current and historic data is stored for the purpose of reporting and analytics. Setting up such a data warehouse and maintaining it is part of the business intelligence (BI) strategy of a company. However, not all companies have such a centralized database, and even when it exists it rarely contains all the information required for a certain analysis. Therefore, data often needs to be collected, cleansed and combined with other sources. Data collection, cleansing and further pre-processing is usually a very time-consuming task, often taking more time than the actual analyses. Therefore, data collection and pre-processing are always the **first steps** of a BA project. Following the data collection and pre-processing the real data science steps begin with descriptive analytics. Moreover, a BA project does **not** end with prescriptive analytics, with generating an (optimal) decision. The decision has to be implemented, which requires various skills, such as knowledge of change management. To summarize, we distinguish the following steps in a BA project:



During descriptive analytics, you get an understanding of the data. You visualize the data and you summarize it using the tool of statistical data analysis. Getting a good understanding is crucial for making the right choices in the consecutive steps.

Following the descriptive analytics, a BA project continues with predictive analytics. A target value is specified which we want to predict. Based on the available data, the parameters of the selected predictive method are determined. We say that the model is **trained** on the data.

Two terms are closely related to BA: **Data science** and **big data.** Data science is a combination of different scientific fields all concerned with extracting knowledge from data, mainly data mining and statistics. The knowledge base of data scientists and business analysts largely overlap. However, the deliverable of BA is improved business performance, whereas data scientists focus more on methods and insights from data.

Big data differentiates itself from regular data sets by the so called **3 V's:**

1. **Volume**
2. **Variety**
3. **Velocity**

A dataset is considered to be 'big data' when the amount of data is too much to be stored in a regular database, when it lacks homogeneous structure (free text instead of well-described fields), and/or when it is only available real-time. Big data requires adapted storage systems and analysis techniques in order to exploit it.
Big date now receives a lot of attention due to the speed at which data is collected these days. As more and more devices and sensors automatically generating data are connected to the internet (the internet of things) again, the amount of stored data doubles approximately every three years. However, most BA projects do not involve big data, but use with relatively small and structured data sets. It might have been the case that such a dataset had its origin in big data from which relevant information has been extracted.

Example: cameras in metro stations are used to surveil passengers. Using image recognition software in the numbers of passengers can be extracted, which can be used as input for a prediction method that forecasts future passenger volumes.

Non-technical overview

The four steps pre-processing, descriptive, predictive and prescriptive analytics, can also be described as follows:

- **Preparing the data set;**
- **Understanding the data set;**
- **Predicting a target value;**
- **Maximizing the target value.**

Most predictive techniques require that you first structure the data. For example, topics can be extracted from text entered on social media or types of objects can be extracted from images. This brings us to a first distinction:

- **Structured:** structured data usually consists of entries (e.g. people) with attributes (e.g., name, income, sex, nationality). The possible value for the attributes are well-defined (e.g., numerical, M/F, standard country codes). Structured data can be represented as a **matrix**: the rows are the entries, the columns the attributes. Structured data comes in different flavors: for example, it can be numerical (e.g., temperature), categorical (e.g., days of the week), binary (e.g., true/false). Depending

on the type of data different algorithms or adaptions of algorithms are used. If we have univariate data, i.e., data with only one attribute, then we can look at the distribution or compare different data sets. For multivariate data, we can study how the different attributes influence each other.

- **Unstructured data:** has no structure. It might be data from cameras, social media sites, text entered in free text fields, etc. counted in bytes, unstructured data is the majority of the data that is stored today, and it is often also big data. When working with unstructured data, the first step is often to extract features to make it structured and therefore suitable as input for an algorithm working with structured data (e.g., images from road-side cameras are used to extract license plates which are then used to analyze the movement of cars). Dealing with unstructured data is an important part of the pre-processing step. Cleansing is another one. Data often contains impossible values of empty fields.

A **final** important pre-processing activity is **feature engineering**, combining attributes or features into new potentially more useful attributes. For example, combining 'day of week' and 'time' can lead to an attribute 'business hours'.

Next, we explore the data in the **descriptive step.** Typical activities are visualization, different statistical techniques such as **hypothesis testing**, and **clustering.** In clustering, you look for data points that are in some mathematical sense close together. Think about clustering individuals based in income, sex, age and family composition for marketing purposes. In the descriptive step, we do **not** focus on a target value (such as sales or number of patients cured). Having a target value is the defining distinction of predictive analytics. Therefore, predictive analytics is also called **supervised learning.** Supervised learning comes in two flavors:

1. **Regression:** we estimate a numerical value. The best-known methods are linear regression and artificial neural networks;
2. **Classification:** the outcome membership of two or more classes, e.g., whether or not somebody will click on an online ad, or vote on one of a number of parties.

Tooling

A multitude of tools exist to assist the data analyst with his or her task. We first make a rough division between **ad hoc** and **routine** tasks. For routine tasks, standardized and often automated procedures exists for the process steps, often involving dedicated and sometimes even tailor-made software. For example:

- For data collection data warehouses exist with connections with operational IT systems;
- For distribution companies' decision support systems exist that compute the optimal rout of delivery trucks, saving many transit hours and petrol

The 'golden view': data-driven governance in the scoring society

The article identifies an upsurge in data-driven forms of what we term 'citizen scoring' – the use of data analytics in government *for the purpose of categorization, assessment and prediction at both individual and population level.*

From data to data scores

The growing collection of data across social life, what has been described as the 'datafication' of society, is now a prominent feature of politics, economics and culture. The technical ability to turn increasing amounts of social activity and human behavior into data points that can be collected and analyzed has simultaneously advanced a power dynamic in need of investigation and critique.

Data-driven scores and classification that combine data from different sources towards calculating risks or outcomes are emerging as a prime means for categorizations. We are predominantly familiar with these practices in the financial sector, most notably in the corm of the credit score, which increasingly relies on an array of digital transaction to inform predictions about the financial responsibility of individuals. A wider range of consumer scores are now being applied across different economic sectors. Sources of data for such scores may include, for example, an analysis of people's mobile phone use, or the creditworthiness of their social friends. People's social activities are thus increasingly incorporated into particular commercial assessments, which points to a growing integration of social and transactional data sets. This practice builds on established experiences in the marketing industry and, more recently, the platform economy, where consumption patterns are predicted based on variety of social, cultural, health and other date.

Whilst perhaps more normalized in financial and commercial industries, the use of data-driven scores has also reached governmental and public services. Much recent attention has focused on the 'social credit score' being developed in Chine, for example, which aims to integrate the rating of citizens' financial creditworthiness with a wide range of social and consumer behavior to assess people's overall trustworthiness and allow, or deny, services accordingly.

In het study of the uses of data and automated processing in the United States, Eubanks points to a rise of a 'regime of data analytics' in public services, detailing, for example, uses of automated welfare eligibility systems and predictive risk models in child protection akin to the kinds of assessments and categorizations we associate with citizen scoring. Although increasing attention is being paid to developments relating to this kind of citizen scoring, little is known about the uses of new data systems, particularly at the local government level where public services are predominantly provided. We lack analyses of how these systems are implemented and used in practice, changes in governance that occur, how trade-offs are negotiated, and how these relate to the questions and concerns expressed by different stakeholder groups across society.

Austerity and public-private partnerships

Policing has become a prominent area of predictive analytics. The research carried out indicates that predictive policing programming predominantly fall in two areas:

1. **Predictive mapping programs;**
2. **Individual risk assessment programs.**

Most forces using predictive policing programs engage in forms of mapping, which are programs that evaluate police data bout past crimes to identify 'hot spots' of high risk on a map. These are supplied by a range of private companies, including HunchLab, IBM, Microsoft, Hitachi and Palantir. A few are also engaging in individual risk assessment programs which predict how people will behave, including whether they are likely to commit or be victims of certain crimes.

From data warehouses to risk assessments

No standard procedures are in place for how data systems are implemented, discusses and audited. Instead, uses of data systems are approached very differently, with some data-sharing leading to the creation of individual risk-scoring, whilst in other contexts this is not practiced and databases serve predominantly as verification tools or to provide population level analytics. This indicates that whist it is broadly accepted that public services planning requires data and analytics, there is not a shared understanding amongst local authorities as to what is appropriate to do with such technologies.

Despite difference in application, data sharing between agencies and different parts of the council is a prominent trend, described as the creation of 'data warehouses' or 'data lakes', that seek to get 'the golden view' of citizens. This refers essentially to integrated databases that gather information about residents and their interactions with public services, across areas such as housing, education, social services, and sometimes also health and policing.

The creation of this 'golden view' of citizen takes several forms and plays out in a broad range of data applications. We use it here as a metaphor to understand data systems as part of a desire to have both additional and more integrated information about populations as well as more granular information about citizens that form the basis of predictions and can drive actions taken.

We see the varied application of data systems, the significance of contextual factors, such as policy agendas relating to austerity, for turning data-driven technologies in public services, and the further intertwining of government and business spheres. This is significant for the ability to engage citizens in consultations and advance public transparency, as well as positioning public sector workers in an empowered position in relation to negotiating these systems as government agencies become locked-in and reliant on external expertise the less they invest in developing their own internal capabilities. Moreover, in the extensive data collection and sharing, and the onus on prediction and risk assessments as a central feature of data systems in public services, concerns about the implication of these for citizen rights and impact of such decision making on different groups and communities have become prominent.

Negotiations and tensions

There are ongoing tensions emerging as local authorities try to respond to problems facing communities, and doing so with less resources driving a need to be 'smarter' and more efficient. In this section, we outline some key themes emerging with regards to transformations and implications of the implementation of data systems in public services.

Citizen rights and harms

The extent of data collection, who gets to see it, and the lack of transparency around its uses were raised and prominent concerns amongst civil society groups, but are tackled very differently by different councils. Although the EU's GDPR addresses some aspects of data sharing and use, detailed requirements are still unclear and many parts of public service provision are exempt from such regulation. Therefore, local authorities are balancing or engaging in a tradeoff between privacy rights and the rights of vulnerable individuals to protection and care. Indeed, the drive to enable a golden view of citizens by linking up all available data sets comes in part from perceived failing of agencies to adequately share and act on information in order to respond to needs and risks, marked by high-profile cases such as the deaths of Baby P, Victoria and Fiona Pilkington following instances of long-term child abuse.

Less discussed and addressed by local authorities are issues of 'bias' or harms, particular how the use of these new systems might negatively affect people's lives. Such concerns have become particularly prominent as data processes often sit behind a veneer of technological 'objectivity'. This was one of the critiques raised most often by civil society groups. They highlighted concerns about the ways in which a data lens particularly targets those on the margins and how these systems impact citizens' rights and opportunities differently. It is not something that the bulk of the population will ever encounter. It is something you only encounter when you are a part of a risk group, a risk population. Such concerns are echoed in research carried out in other countries that have highlighted how systems like this, which disproportionately draw on and use data about people who make use of social services, are biased through the over-representation of a particular part of the population. The variables being used can in practices be proxies for poverty, for example by using the length of time someone has been on benefits as a variable influencing risk assessments.

Related to this, several civil societies raised the issue of stigmatization as a central feature of citizen scoring, highlighting how the creation of data warehouses, risk assessments and predictions in itself can be harmful: because of this kind of quantification and categorization approach that data analytics actually demands and the use of ever more sensitive data, there are people who feel sidelines, maligned, judged and stereotyped. Further, none of the case studies we analyzed included a means for people who had been scored to know their score, how it was generated and how to interrogate it. This inability to see of talk back was seen as having significant democratic implications in terms of due process and can lead to differential treatment and opportunity given the way that someone may unknowingly be affected by a score.

Experiences amongst service users and communities point to the need to engage more comprehensively with the day data systems relate to different activity that might lead to a range of harms and feelings of being targeted. This requires a re-evaluation of how authorities and the state might be perceived as not necessarily benign, and that technologies are not necessarily neutral. Whilst harmful outcomes relating to data collection and use might not be intentional, such evaluations point to the need to consider how data has the potential to facilitate punitive measures. Yet what kind of impact would need to be assessed and how evaluations on actions taken on citizen scores would be carried out remain difficult areas as there is no clear line of accountability for any one system that is distributed across different people and uses. Moreover, councils pointed to a lack of recourses in pursuing any comprehensive evaluations or impact assessments of transformations in practices and provision with the implementation of new data systems.

Professional authority and operational logics

This question of how to evaluate or assess impacts gains further pertinence as the tensions and negotiations surrounding the harms and rights infringements that may arise with the use of data systems in public services are simultaneously playing out in a context of changing practices and organizational transformations that position different understandings and activities at odds. In building a culture of data collection, we found a concern amongst both civil society groups and frontline staff about a fundamental re-orientation of professional practices and routines, relationships and the kinds of information deemed valuable in delivering public services. In determining a family's needs, for example, a member of professional association for social worked noted that 'the systems are set up for social workers to collect data as performance management', pointing to a concern that this 'can divert the social worker from being able to understand the case because the sort of data that they're collecting, they might be lost in there, the complexities of the case'.

In the prominent application of data systems for the purpose of identifying and measuring risks, we are also confronted with a general shift within public administration towards risk management as a new 'paradigm' of operations. The way in which this shifts authority from public sector workers themselves towards computational outputs was a recognized tension and frequently addressed through an explicit emphasis on professional judgement as the central pillar for any decision-making, regardless of the implementation of data systems.

We see how at the level of management and development of data systems in the context of public services challenges are predominantly seen as either technical and cultural in nature. Issues pertaining to data quality or organizational skepticism towards technology are current obstacles, but are of a kind that can eventually be overcome through 'better' data practices that ultimately fit a shift towards data-driven governance. This understanding of challenges marks a significant discrepancy with the more fundamental concerns expresses by stakeholder groups from both civil society and frontline staff. Here we see a concern with social and political issues that speak to tensions at the core of what the 'golden view' of citizens might mean, in terms of different harms, rights, and the potential for enacting agency both as service users and professionals.

Transformations in governance: deconstructing the 'golden view'

The turn to data-driven technologies raise concerns across different stakeholders not just about the lack of transparency and likelihood of errors and bias in the design and use of these systems, but also about a more fundamental shift in what constitutes or is privileged as social knowledge, the kinds of actions that might be taken on such knowledge, and the way in which this positions citizen as subjects of governance.

Concerns point to the implications of 'seeing' people through data within this context, and the abstracted and reduced understanding this may lead to when relied upon the expense of other types of knowledge. In conjunction with the deskilling and disempowerment of professionals as the use of data systems grows, issues raised by stakeholders speak to a perceived danger that the messiness of people and lived experiences is necessarily sidelined or ignored for the algorithmic processing of information. With the turn to algorithmic decision-making in governance, authority and expertise is transferred to calculative devices seeing to capture risk over and above other forms of expertise. As such, citizens are positioned in the 'golden view' not as participants or co-creators, but primarily as (potential) risks, unable to engage with of challenge decisions that govern their lives.

Moreover, concerns with targeting and stigmatization, particularly of marginalized and poor groups in society, highlight the way these systems attribute risk factors to individual's' behavior and characteristics, shifting the burden of responsibility for social ills onto individuals over and above solutions. When the focus is on individuals, predicting risks of committing crime through data-driving profiling, for instance, is comfortable presented as a 'solution' for tackling increasing crime levels whilst doing little to engage with an underlying cause of crime. These systems, in their emphasis on correlation over causation, can individualize social problems by directing attention away from structural causes of social problems.

Conclusion

The introduction of predictive analytics, scoring systems, intelligent databases and data warehouses into local government is a rapidly emerging feature of datafication. For public services, these systems are said to offer an opportunity to allocate resources and respond to needs more effectively. However, little is known about the kinds of systems in place, how and what they are used, and what practitioners and stakeholders think about these developments. This is especially a challenge in what we have identified as both a regulatory and interpretive vacuum that signifies a lack of shared understanding of not only what constitutes data-driven decision-making and algorithmic processing of information, but also what is appropriate to do with such systems.

The turn to scoring systems and predictive analytics is being fueled by an austerity context in which councils have faces substantial cuts. While these technologies are being implemented as 'smart' and effective solutions for better service provisions, they are introduced in the context of service reduction. Further, we can observe a strong reliance on commercial systems that provide additional challenges to transparency and incorporate a wider set of (transactional, social, etc.) data on people into public sector decision-making. In this setting, shifts in organizational practices and logics that implicate the role of professional judgment and the extent to which data systems come to guide decision-making have led to prominent

concerns amongst stakeholder's groups in civil society that are not necessarily considered within local authorities and partner agencies. These include concerns beyond questions of transparency, bias and discrimination, and point to broader worries about targeting and stigmatization, and how people come to be 'seen' and engages with as citizens and service users.

WEEK 2 - Privacy and data protection law

Mireille Hildebrandt – Law for computer scientist and other folk

Chapter 5

Privacy and Data Protection

5.1.2 From liberty rights to social, economic and further rights

Humans rights law was originally focused on the protection of individual citizens against powerful states. We call these rights first generation human rights, and they are best described as the **subjective right** that the state refrains from interference with the legal good that is protected by such rights. This is why they are often called **liberty rights.**

These legal goods are: privacy, non-discrimination, bodily integrity, freedom of movement, the presumption of innocence, a fair trial, freedom of expression, freedom of association, freedom of religion and voting rights. Not that these legal goods are considered worthy of protection as public goods, because a society that does not protect them cannot support a viable democracy that depends on independence of thought and unhindered development of both individual and group identities. The focus is on public goods that protect individual persons as autonomous agents in a democratic polity and on negative obligation of the state towards its citizens.

5.2 The concept of Privacy

Before investigating the right to privacy, we will first inquire into the nature of privacy itself. The reason is that computer science has a specific relationship with privacy, notably in the context of digital security and cryptography. In that context, privacy is often seen as a subset of security, focused on hiding or removing the link between data and whoever the data refers to, or on encrypting the data to safeguard confidential data against eavesdropping. This has, as a consequence, meant that privacy protection is restricted to:

1. Anonymization or pseudonimisering of personal data, by way of deleting or separating identifiers and to
2. Hiding the content by means of encryption or other security measures.

Consider the following data points:

- Your name;
- Your bank account;
- The taxes your mother pays
- What kind of socks you wear;
- The logs of your surfing behaviour on the net etc.

Should we qualify this data as part of the privacy of the person the data refers to? To answer this question, we need to check what falls within the value of, the interest in, or the right to privacy:

- When (under what conditions)?
- With regard to whom (is data on my mother part of my privacy)?
- Where (are specific locations more privacy sensitive than others)?
- For what reason (what could make my socks relevant to my privacy)?

The American privacy scholar and lawyer Daniel Solove made an insightful attempt to approximate the concept of privacy in terms of six categories that are partly overlapping, while thus covering much of what we intent when referring to privacy:

1. The right to be left alone;
2. Limited access to self;
3. Secrecy – concealment;
4. Control over personal information;
5. Personhood – protection of identity, dignity; and
6. Intimacy

5.2.2 Privacy and technology

In a famous article in the Harvard Law Review, US legal scholars Samuel Warren and Louis Brandeis discusses the need to protect oneself against publication of photographs without permission, to enable social withdrawal. In that article, they formulated the right to privacy as the right to be left alone, basically arguing for the existence of privacy tort whenever this right was infringed upon without justification. When Brandeis later served as justice in the Supreme Court, he argued that such a right to be left alone must be 'read into' the US Constitution, notably into the Bill of Rights, thus vouching for a right to privacy against the state.

The concept of informational privacy, as control over information, informs much of the debate about privacy and data protection in our current age. It is interesting to note that it emerged in counterpoint to the rise of databases in public administration, as well as private enterprise. The fact that data was collected, sorted, and recorded, enabling retrieval as well as aggregation, gave rise to new types of transparency, and new types of threats to personal identity. This was related to the fact that in this era the data collected and stored was mostly stable data, allowing the mapping of both individuals and populations in consistent and foreseeable way, without the kind of dynamic and unstructured big data capture that characterizes the current era.

After the rise of the internet combined with the capture of big data and data-driven techniques to infer new information, the need for a more complex and contextual right to privacy seems obvious. Negative freedom will not do, as data abounds and is captured beyond one's control on a permanent basis. For the same reason, positive freedom seems unattainable, as consent loses its meaning amidst the volume, variety and velocity of data capture, storage and use. A more practical and effective way of understanding privacy should therefore combine negative and positive freedom, while highlighting the relationship with identity-construction, not merely identification.

The definition of Agre and Rotenberg may be the most apt for the era of proactive and pre-emptive computing infrastructure, depicting the right to privacy as:
The right to be free of unreasonable constraint on the building of one's identity.

5.3 The right to privacy

Privacy is a value, an interest, a right, or a good. It can be analyzed from an ethical perspective (as a value, virtue or duty), from an economic perspective (as a utility, a preference of an interest) and from the perspective of political theory (as a public and private good).

5.3.1 The right to privacy: constitutional law

The right to privacy is a subjective right, attributed by objective law. The most obvious branch of objective law that attributes the subjective right of privacy is constitutional law, which often contains a section that aims to protect citizens against overly invasive powers of state. The industrial revolution of the nineteenth century gave rise to powerful economic actors whose ability to infringe privacy, freedom of information, and non-discrimination increasingly matched the powers of state. This has led courts to recognize a so-called **horizontal effect** of constitutional rights such as privacy. This entails that protection against such infringements is a duty of the state, meaning that citizens can sue the state for failing to impose prohibitions to infringe these rights upon powerful players in the private sector. This is called **indirect horizontal effect,** because it cannot be invoked directly against private parties.

5.3.2 The right to privacy: international law

Protection of human rights requires a resilient system of checks and balances, that is, a series of institutional safeguards to ensure that the state does not claim unreasonable expectations and faces a stringently independent judiciary to keep the powers of the state 'in check'. The need to protect subjects of the state against the state, gave rise to international human rights law, which provides an extra layer of checks and balances. Privacy is explicitly protected by article 17 of the United Nations, International Covenant on Civil and Political Rights of 1966, and by Article 8 ECHR.

5.3.4 Article 8 ECHR

The right to privacy that is articulated in Article 8 ECHR is not only relevant for bodily integrity, decisional privacy, and the other aspects of privacy, but also directly affects issues of cybercrime and copyright. This is due to the fact that cybercrimes may violate privacy (hacking, data breaches), or that copyright holders may violate privacy when disseminating their works

(photographs, texts), but also because the investigative measures that aim to detect cybercrime and violations of copyright often infringe upon the right to privacy as protected in Article 8.

Article 8 consists of two paragraphs. The first paragraph concerns the question of whether privacy is infringed, the second paragraph clarifies under what conditions an infringement is justified.

1. *Everyone has the right to respect for his private and family life, his home and his correspondence*

The legal effect generated by this paragraph is 'an infringement of privacy', and this infringement depends on the following **alternative** legal conditions:

- Private life is not respected;
- Family life is not respected;
- The protection of one's home is not respected; and
- The confidentiality of one's correspondence is not respected.

The ECtHR takes the view that these concepts require a broad rather than a narrow interpretation, bringing a wide variety of situations, events, relationships, and contexts under the protection of Article 8.

Private life can be at stake in the context of work, meaning that a search of an office space may be an infringement of privacy. Family life is at stake when a state prohibits members of a family living together, for instance in the case of a refusal to provide a residence permit for a partner from another state, or of a parent wishing to further develop a relationship with their child despite not being married to the other parent. Protection of the home may become relevant when a person has taken residence in a house they neither own nor rent, meaning that the need to respect one's home is not dependent upon ownership or contract. The confidentiality of communication has been interpreted to include letters, telephone calls, and more recently all types of internet-enabled communication that is not public. Privacy, as protected by Article 8, clearly concerns physical, spatial, contextual, decisional, communicative, and informational privacy, and although Article 8 addresses the contracting states, the indirect horizontal effect has been recognized by the ECtHR, requiring states to ensure proper protection against

violations by others than the state. Note that the individual complaint right to the ECHR can **only** be invoked against a state, not against a company. To invoke direct horizontal effect, a person needs to sure the tortfeasor in national court.

Note: an infringement of privacy is **not** the same as a violation of the right to privacy. Once the legal effect of an infringement has been established by the ECtHR, it will investigate whether the state has a valid justification.

2. *There shall be **no** interference by a public authority with the exercise of this right **except** such as in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*

The legal effect of a **valid justification** is that, despite the infringement, Article 8 is **not** violated. This effect depends on the following **cumulative** legal conditions:

- The infringements have one of more of the following legitimate aims:
    1. National security, public safety, the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others;
    2. The infringements is in accordance with the law; and
    3. The infringement is necessary in a democratic society;

The second paragraph of Article 8 thus requires a **triple test**, meaning that all three conditions **must** be met. These conditions are often summed up by stating that any infringing measures by the state must:

1. Have a legitimate aim;
2. Have a basis in law; and
3. Be proportional in relation to the aim served.

The Court has developed another **triple test** to decide whether an infringement has a **proper basis in law** (see point 2):

1. The legal competence to take infringing measures must be **accessible**, knowable for citizens to whom it will apply;
2. The infringement must be **foreseeable,** which means sufficiently specified; and
3. The quality of the law must include **sufficient safeguards** that limit the exercise of the competence in time and space, specifying the extent to which privacy may be infringed, and notably requiring independent oversight (e.g. warrants) in the case of more serious infringements.

If privacy is infringed with a legitimate aim, based on a legal competence that is accessible, foreseeable, while having sufficient safeguards, the final test is the **proportionality test.**

The proportionality test entails that the ECtHR investigates whether the measure was necessary in a democratic state, which requires a **pressing social need** to resort to such measures.

- Under this criterion the Court will examine the gravity, invasiveness, and seriousness of the infringement in relation to the importance and seriousness of the aim served.
- This criterion basically requires that the measures taken can be reasonably be expected to be effected, because a measure that is not effective cannot be necessary.
- The proportionality test includes a **subsidiarity test;** if another measure which is less infringing is feasible or sufficiently effective, the measure is **not** proportional.

5.3.5 Case law Article 8 ECHR regarding surveillance

When developing computing architectures, whether in the context of databases, streaming data, machine to machine communication, knowledge discovery in databases, machine learning, or cryptographic infrastructures, computer scientists lay the foundations for the ICIs that enable the processing, storage, interlinking and inferencing of behavioural and other personal data. This may regard online clickstream behaviour, location, and mobility data. Governments, tasked with the investigation and prosecution of criminal offences and the protection of national and public security, have many incentives to gain access to such data. Apart from the struggle against serious crime and threats to national security, governments

need to collect taxes, attribute social benefits, take precautionary measures regarding public health, and safeguard the economic welfare of the country. All these tasks fall within the scope of the legitimate aims enumerated in Article 8 paragraph 2 ECHR. This raises the question under what conditions surveillance measures can be qualified as 'in accordance with the law' and if so, when they are consider proportional to the targeted aim.

To understand how the Court deals with various types of electronic surveillance, we will discuss two cases of post-crime surveillance and two cases of pre-crime surveillance.

5.3.5.1 Post-crime surveillance

In **Malone vs UK** the ECtHR determined that the UK was in breach of Article 8 ECHR, where it allowed the interception of telephone conversations by the police upon a warrant issued by the Secretary of State. The Court determined that for such a measure to be 'in accordance with the law', it must **not** merely have a basis in domestic law (meaning a legal power), but must also be foreseeable and sufficiently limited as required by the rule of law:

ECHR: since the implementation in practice of measures of secret surveillance of communication is not open to scrutiny by the individuals concerned or the public at large, it would be contrary to the rule of law for the legal discretion granted to the executive to be expressed in terms of an unfettered power. Consequently, the law must indicate the scope of any such secretion conferred on the competent authorities and the manner of its exercise with sufficient clarity.

When applying this interpretation, the Court finds that:

'The foregoing considerations disclose that, at the very least, in its present state the law in England and Wales governing interception of communications for police purposes is somewhat obscure and open to differing interpretations. The Court is, however, required under the Convention to determine whether, for the purposes of paragraph 2 of Article 8, the relevant law lays down with reasonable clarity the essential elements of the authorities' powers in this domain. In the opinion of the Court, the law of England and Wales does ***not*** *indicate with reasonable clarity the scope and manner of exercise of the relevant discretion conferred on the public authorities. To that extent, the minimum degree of legal protection to which citizens are entitled under the rule of law in a democratic society is lacking'.*

In this case, Malone not only claimed that the interception of the content of his telephone conversations violated his right to privacy under the Convention, but also that the capture of what we would now call metadata violated his right. The Court states, with regard to 'metering':

'The process known as 'metering' involves the use of a device (a meter check printer) which registers the numbers dialled on a particular telephone and the time and duration of each call. In making such records, the Post Office, makes use only of signals sent to itself as the provider of the telephone service and does not monitor or intercept telephone conversations at all. From this, the Government drew the conclusion that metering, in contract to interception of communications, does ***not*** *entail interference with any right guaranteed by Article 8. On the evidence adduced before the Court, apart from the simple absence of prohibition, there would appear to be no legal rules concerning scope and manner of exercise of the discretion enjoyed by the public authorities. Consequently, although lawful in terms of domestic law, the interference resulting from the existence of the practice in question was* ***not*** *'in accordance with the law' within the meaning of paragraph 2'.*

Note: the ECtHR established that the practice of 'metering' is lawful under UK law, but in violation of Article 8.2 ECHR. Both the interception and the metering violate Article 8.2 ECHR because they are **not** 'in accordance with the law' as required by a treaty that binds the UK. This means that the UK has violated its legal obligation under the Convention and is now bound to ensure that these types of surveillance measures are based on a domestic law that both constitutes and sufficiently restricts its legal powers.

In **Huvig & Kruslin vs France** the ECtHR determined that Article 8 was breached. The Court extensively refers to its contentions in the Malone judgement as to the requirement of such interception being 'in accordance with the law'. It then states:

'*the system does not for the time being afford adequate safeguards against various possible abuses. For example, nothing obliges a judge to set a limit on the duration of telephone tapping. In short, French law, written and unwritten, does* ***not*** *indicate with reasonable clarity the scope and manner of exercise of the relevant discretion conferred on the public authorities. This was truer still at the material time, so that Mr Kruslin did not enjoy the minimum degree*

of protection to which citizens are entitled under the rule of law in a democratic society. There has been a breach of Article 8'.

5.3.5.2 Pre-crime surveillance (including surveillance by the intelligence services)

In **Klass vs Germany**, the ECtHR decided a case regarding surveillance measures taken by the secret services in Germany. The following will clarify how the Court argues points of law and thus shapes the interpretation of legal conditions:

'*All applicants claim that the Articles are contrary to the Convention. They do not dispute that the State has the right to have recourse to the surveillance measures contemplated by the legislation; they challenge this legislation in that it permits those measures without obliging the authorities in every case to notify the persons concerned after the event, and in that is excludes any remedy before the court against the ordering and execution of such measures'.*

The Court first discusses the admissibility of the complaint, raising the question whether the applicant is a victim of violation by on of the Member States.

'*The Court points out that where a State institutes secret surveillance the existence of which remains unknown to the persons being controlled, with the effect that the surveillance remains unchallengeable, Article 8 could to a large extent be reduced to a nullity. It is possible in such a situation for an individual to be treated in a manner contrary to Article 8, or even to be deprived of the right granted by that Article, without his being aware of it and therefore without being able to obtain a remedy either at the national level or before the Convention institutions. The Court finds it unacceptable that the assurance of the enjoyment of a right guaranteed by the Convention could be thus removed by the simple fact that the person concerned is kept unaware of its violation. Having regard to the specific circumstances of the present case, the Court concludes that each of the applicant is entitled to '(claim) to be the victim of a violation' of the Convention, even though he is not able to allege in support of his application that he has been subject to a concrete measure of surveillance'.*

This entails that the Court makes an exception to the requirement that applicants must claim and demonstrate to be a victim of violation in concrete terms.

The Court begins by investigating whether the legislation that is contested by the applicants, constitutes an **interference** with Article 8.1 ECHR. As is often the case, the Court takes a broad view of the scope of the first paragraph and decides that the legislation constitutes an infringement. The next question is whether the infringement is **justified.** The Court first tests whether the infringement is **in accordance with the law:**

*'In order for the 'interference' established above not to infringe Article 8, it **must,** according to paragraph 2, first of all have been in accordance with the law. This requirement is fulfilled in the present case since the 'interference' results from Acts passed by Parliament, including one Act which was modified by the Federal Constitutional Court, in the exercise of its jurisdiction, by its judgment of 15 December 1970. In addition, the Court observes that, as both the Government and the Commission pointed out, any individual measure of surveillance has to comply with the strict conditions and procedures laid down in the legislation itself'.*

This leads the Court to test whether the interference has a **legitimate aim:**

***'**The G 10 defines precisely, and thereby limits, the purposes for which the restrictive measures may be imposed. It provides that, in order to protect against 'imminent dangers' threatening 'the free democratic constitutional order', 'the existence or security of the Federation or of a Land', 'the security of the (allied) armed forces' stationed on the territory of the Republic or the security of 'the troops of one of the Three Powers stationed in the Land of Berlin', the responsible authorities may authorise the restrictions referred to above. The Court, sharing the view of the Government and the Commission, finds that the aim of G 10 is indeed to safeguard national security and/or to prevent disorder or crime in pursuance of Article 8 paragraph 2. In these circumstances, the Court does not deem it necessary to decide whether the further purposes cited by the Government are also relevant'.*

This brings the Court to test the **final criterion** of the triple test, investigating whether the interference is necessary in a democratic society. Below you will find an extensive quotation

of (part) of the reasoning of the Court regarding the question whether the interference enabled by the legislation is **proportional,** considering what is at stake.

'As concerns the fixing of the conditions under which the system of surveillance is to be operated, the Court points out that the domestic legislature enjoys a certain discretion. It is certainly not for the Court to substitute for the assessment of the national authorities any other assessment of what might be the best policy in this field. Nevertheless, the Court stresses that this does not mean that the Contracting States enjoy an unlimited discretion to subject persons within their jurisdiction to secret surveillance. The Court, being aware of the danger such a law poses of undermining or even destroying democracy on the ground of defending it, affirms that the Contracting States may not, in the name of the struggle against espionage and terrorism, adopt whatever measures they deem appropriate.

The Government maintain that Article 8 paragraph 2 does not require judicial control of secret surveillance and that the system of review established under the G 10 does effectively protect the rights of the individual. The applicants, on the other hand, qualify this system as a 'form of political control', inadequate in comparison with the principle of judicial control which ought to prevail. It therefore has to be determined whether the procedures for supervising the ordering and implementation of the restrictive measures are such as to keep the interference resulting from the contested legislation to what is 'necessary in a democratic society.

Review of surveillance may intervene at three stages: when the surveillance is first ordered, while it is being carried out, or after is has been terminated. As regards the first two stages, the very nature and logic of secret surveillance dictate that not only the surveillance itself but also the accompanying review should be effected without the individuals' knowledge. Consequently, since the individual will necessarily be prevented from seeking an effective remedy of his own accord or from taking a direct part in any review proceedings, it is essentials that the procedures established should themselves provide adequate and equivalent guaranteed safeguarding the individuals' rights.

In addition, the values of a democratic society must be followed as faithfully as possible in the supervisory procedures if the bounds of necessity, within the meaning of Article 8-2, are not to be exceeded.

One of the fundamental principles of a democratic society is the rule of law, which is expressly referred to in the Preamble to the Convention. The rule of law implies, inter alia, that an interference by the executive authorities with an individual's right should be subject to an effective control which should normally be assured by the judiciary, at least in the last resort, judicial control offering the best guarantees of independence, impartiality and a proper procedure.

The Court considers that, in a field where abuse is potentially so easy in individual cases and could have such harmful consequences for democratic society as a whole, it is in principle desirable to entrust supervisory control to a judge.

Nevertheless, having regard to the nature of the supervisory and other safeguards provided by G 10, the Court concludes that the exclusion of judicial control does not exceed the limits of what may deemed necessary in a democratic society.

In the opinion of the Court, it has to be ascertained whether it is even feasible in practice to require subsequent notification in all cases.

Subsequent notification to each individual affected by a suspended measure might well jeopardise the long-term purpose that originally prompted the surveillance. Furthermore, as the Federal Constitutional Court rightly observed, such notification might serve to reveal the working methods an fields of operation of the intelligence services and even possibly to identify their agents.

In the Court's view, in so far as the 'interference' resulting from the contested legislation is in principle justified under Article 8-2, the fact of not informing the individual once surveillance has ceases cannot itself be incompatible with the provision since it is this very fact which ensures the efficacy of the 'interference'.

This particular case is a landmark case that functions as a building block for reasoning in similar cases and requires the contracting states to incorporate necessary safeguards when developing and implementing legislation that enables surveillance by intelligence agencies.

In 2006, the ECtHR decided the case of **Weber & Saravia vs Germany**, once again testing legislation regarding so-called 'strategic monitoring' by intelligence services. In this case, the Court specifies in more detail what qualifies as 'interferences' that are 'in accordance with the law'. Although, after having conducted the triple test, the Court decided that the contested legislation did not violate Article 8 ECHR. I will quote the legal conditions summed up by the Court to attain the legal effect of such interferences qualifying as being 'in accordance with the law':

'*In its case-law on secret measures of surveillance, the Court has developed the following minimum safeguards that should be set out in statute law in order to avoid abuses of power:*

- *The nature of the offences which may given rise to an interception order;*
- *A definition of the categories of people liable to have their telephones tapped;*
- *A limit on the duration of telephone tapping;*
- *The procedure to be followed for examining, using and storing the data obtained;*
- *The precautions to be taken when communicating the data to other parties; and*
- *The circumstances in which recordings may or must be erased or the tapes destroyed.*

5.4 Privacy and Data Protection

Since the CFREU (or 'the Charter') has been in force, the EU 'has' two fundamental rights regarding the processing of personal data:

- **Article 7 respect for private and family life**: everyone has the right to respect for his or her private and family life, home and communications;
- **Article 8 – Protection of personal data**

Article 52 of the Charter clarifies the relationship between Article 7 of the Charter and Article 8 ECHR, which both refer to the right of privacy:

In so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection.

This stipulates that Article 7 from the Charter **cannot** be interpreted as providing less protection compared to Article 8 ECHR, but may be interpreted as attributing additional protection. To the extent that Article 8 CFREU corresponds to Article 8 ECHR, it can – similarly – not be interpreted as providing less protection than Article 8 ECHR, but it may provide additional protection.

5.4.1 Defaults: an opacity right and a transparency right

Some authors have argued that whereas, by default, the right to privacy is foremost an **opacity** right, data protection is foremost a **transparency right.** As an opacity right, the right to privacy aims to safeguard a private sphere for individual citizens, where they can basically ward-off interference by others, most notably the state. This highlights the idea that privacy is a **liberty right,** a negative right that obligates others to refrain from interference with the good that is protected. As a transparency right, the right to data protection aims to ensure that whenever personal data is processed (which included collection, access, manipulation, and any other usage) such processing must be done in a transparent manner, in compliance with a set of conditions which should ensure fair and lawful processing.

Note that the opacity concerns the private sphere of an individual person, whereas the transparency concerns the state and other powerful actors when processing personal data. This accords with the core tenets of the Rule of Law, which hold that whereas government should be as transparent as possible, citizens should be shielded from intrusive transparency by the government.

Also, as discussed above, even though privacy is an opacity right that requires the state to refrain from interference (negative freedom) the right to privacy may, nevertheless, impose positive obligations on the state to enable individuals to exercise their right. Similarly, though data protection is a transparency right that should enable individuals as well as others to act on their personal data (positive freedom), while imposing a number of positive obligations on those who determine the purpose of processing, the right to data protection may, nevertheless, require that others abstain from processing personal data, thus imposing negative obligations on them.

5..4.2 Distinctive but overlapping rights: a Venn diagram

Though one may be tempted to see the right to data protection as a subset of the right to privacy, this is not correct. Within the context of the EU, the right to privacy entails both more and less than the right to data protection.

Whenever the processing of personal data constitutes an interference with the right to privacy, there is an overlap. The right to privacy, however, also concerns interference with bodily integrity, decisional privacy, privacy of the home, and correspondence when no processing of personal data is involved. This is where the right to privacy entails **more** than the right to data protection.

Similarly, the right to data protection also concerns the processing of personal data when there is no interference with the right to privacy, for instance, when one's personal data are processes on one's own request, for example, the processing of an address or banking details to deliver goods and charge one's account as a consequence of the sale of a book.

Note that if such data are subsequently used for other purposes, for example, to support the business model of a web shop by way of targeted advertising, privacy may be at stake. Whether or not this is the case also relates to the fact that the right to privacy, as discussed above, is primarily at stake in the vertical relationship between a government and its citizens, whereas the right to data protection seems to be applicable to all those who process personal data. This is certainly the case for data processing that falls under the scope of the GDPR.

5.5.1 EU and US data protection law

In the US, data protection is part of the right to privacy (in Constitutional and tort law) and subject to sectorial legislation, notably with regard to finance, healthcare, special protection of children, and consumer protection. There is no general law on data protection, apart from the 1974 Privacy Act (which only applies to Federal Agencies). This means that the protection of personal data varies with the context of processing. In commercial contexts, much of the actual protection depends on the competences of the Federal Trade Commission, based on section 5 of the FTC Act.

Based on this, the FTC is tasked with protecting consumer privacy and data security in commercial context. The notion of a reasonable expectation of privacy is a core concept, because consumer trust is pivotal for a well-functioning marked in ecommerce. The FTC deals with violations on a case-by-case basis, but also issues so-called 'rulings' if it believes specific types of violations are prevalent. Such 'rulings' basically declare how the FTC will use its Article 5 competence, thus encouraging companies to change their behaviour. The FTC is often qualified as 'the regulator' concerning informational privacy, due to its central role in US policy-making regarding data protection.

In the EU, the situation is altogether **different**, due to the general applicability of EU data protection law, which does not depend on whether a violation can be framed as 'an unfair or deceptive act in or affecting commerce'.

One could say that whereas in the US the processing of personal information is allowed unless is has been explicitly restricted, in the EU any processing of any personal data in any context is conditioned by a set of rules and principles that impose obligations on those who process data and attribute rights to those whose personal data is at stake.

5.5.2 EU data protection law

The GDPR protects the fundamental right to data protection as stipulated in Article 8 in the Charter. However, the GDPR goes beyond this, explicitly aiming to protect **all the fundamental**

rights and freedoms that are implicated by the processing of personal data. But this is not only goal of the Regulation. At the same time, the Regulation aims to prevent that different levels of data protection within the jurisdiction of the Member States result in obstructions of the internal market. So, harmonization of protection to ensure a free flow of personal data is the second, equally important goal of the GDPR.

As paragraph 3 of Article 1 GDPR clarifies, this Regulation involves 'full harmonisation', which means that Member States are not allowed to provide either less or more protection than what is offered in the Regulation (with the exception of explicitly formulated discretion). **Full harmonization** ensures the absence of obstruction of the internal market due to different requirements in terms of data protection. The fact that the GDPR is a regulation instead of a directive confirms the wish to eradicate such obstructions, thus hoping to boost data-driven business across national borders.

5.5.2.2 Material and territorial scope

The material scope of the GDPR is limited to 'the processing of personal data' (Article 2.1). The definition of 'processing', however, is very broad, as Article 4 (2) reads:

'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaption or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction'.

The GDPR does **not** apply to the processing of personal data within the context of a household and it does not apply to processing of personal data in the context of the prevention and prosecution of crime and threats to public security. The household exception will usually exempt the users of social networks, but not the providers.

Next to exemptions of Article 2, Article 33 states that Member States may enact legislation to restrict the applicability of specific GDPR provisions, if they regard measures that are necessary in a democratic society, targeting a limited set of goals, such as national security, defence, public security, the prevention, investigation, detection and prosecution of criminal offences, or breaches of ethics for regulated professions, an important object of general public interest of a Member State or of the EU, including monetary, budgetary, and taxation matters. Note that though restrictions based on these goals are allowed if they pass the proportionality test ('necessary in a democratic society' clearly refers to Article 8.2 ECHR), they also require a basis in law. Any such restrictions are only valid insofar as they respect the essence of the fundamental rights and freedoms.

The territorial scope of the GDPR is defined in Article 3:

1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.

If a company decides to offer goods or services (whether or not they are free) that involve the processing of personal data of data subjects in the Union, or monitor their behaviour in the Union, the GDPR applies, irrespective of whether the company is established in the Union. Consider that this jurisdiction is limited to data subjects who are in the Union; it does not apply to EU citizens outside the Union, though it does apply to non-EU citizens when they are in the Union.

5.5.2.3 Personal data and data subject

Article 4 GDPR clarifies that:

'personal data' means:

- Any information;
- Relating to;
- An identified or;
- Identifiable natural person ('data subject').

Where ‘an identifiable person’ is defined as:

- One who can be identified;
- Directly or indirectly;
- An identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity.

So, at some point data about the weather, about room temperature, about the arrival of a train may become personal data, when it can be related to a person that can be singled out. Recital 26 reads:

‘*To determine whether a natural person is identifiable, is that it is* ***reasonably likely*** *that a person can, for example, be singled out. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments’.*

Here we see that the **reasonably likely** criterion should be understood as an objective criterion, taking into account the costs, the time, and effort, and the available technical means at the time of processing.

In the case of **Breyer vs Germany** the CJEU decided that even a dynamic IP address may qualify as a personal data, depending on whether the link with a specific person can be made. The case concerned government websites that processed dynamic IP addresses, keeping them longer than was necessary for providing access to the sites. What made this case special is that the ability to link the IP address to a specific person was not in the hands of the operators of the government website but in the hand of internet service providers (ISPs). The CJEU found that because ISPs could be ordered by a court to provide information about the user of a dynamic IP addresses, this IP address should **not** be considered anonymous.

So, personal data is **any data** that relates to an **identifiable natural person** (excluding legal persons such as corporations) and a **data subject** is the identifiable natural person to whom the data relate.

The material scope of the GDPR regards the processing of personal data. This implies that to avoid applicability of the GDPR, one could 'simply' anonymize previously personal data. There are two caveats here:

1. Anonymization is itself a form of processing, and thereby requires a valid legal ground;
2. Anonymization is not easy, because the risk of re-identification easily turns 'anonymous' data into identifiable and thus personal data. In practice, anonymization will often remove so much information from data that it is no longer relevant for the purpose of processing. To better understand the difference between personal and anonymized data, we an best check the definition of **pseudonymization** of Article 4(5):

- *the processing of personal data in such a manner that*
- *the personal data can no longer be attributed to a specific data subject*
- *without use of additional information*
- *provided that such additional information is kept separately and*
- *is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified of identifiable natural person.*

First, we see that pseudonymous data is defined as a subset of personal data. **Second,** it is defined as data from which any identifying information has been removed and stored separately, subject to technical and organizational measures that resist re-identification.

Pseudonymisation is a way to comply with data protection law (by ways of data minimization), **not** a way to avoid applicability. With regard to encryption, key management that enables a party other than the data subject to decrypt, will mostly qualify as pseudonymisation, **not** as anonymization. Bear in mind that the definition of pseudonymisation in the GDPR defines the

condition of the relevant legal effect, irrespective of how other disciplines define pseudonymisation.

5.5.2.4 Data controller and data processor

Article 4 (7) GDPR defines '**controller'** as:

- the natural or legal person, public authority, agency or any other body
- which alone or jointly with others
- determines the purposes and means of the processing of personal data.

The definition of 'data controller' is crucial, because the 'data controller' is both accountable and liable for compliance with **all the obligations** of the GDPR, including obligations to implement a proactive approach to potential risks to the fundamental rights and freedoms of data subjects. The 'data controller' is basically defined as whoever determines the purpose of processing, whereby the CJEU checks who determines such purpose in practice, **not merely** on paper. The 'data controller' also determines the means of processing, but this can be outsourcing to a data processor as (Article 4 (8)):

- a natural or legal person, public authority, agency or any other body;
- which processes personal data **on behalf of the controller**

Here, we clearly see that the data controller remains accountable for the choice of the means of processing, even is that choice is made by a processor.

5.5.2.5 Legal ground for lawful processing of personal data

The processing of personal data is only allowed on the basis of one of six legal grounds. Article 6 GDPR reads:

a. the data subject has given **consent** to the processing of his or her personal data for one or more specific purposes → under the GDPR it is explicitly clear that consent is only valid if the purpose has been specified. As Article 5 stipulates that data may only processed if necessary for the specified purpose, this means that consent can only concern the processing of personal data that is necessary for the purpose that was communicated. All the other grounds stipulate that the processing must be necessary in relation to the ground;

b. processing is necessary for the **performance of a contract** to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract → once the contract has been concluded and performed and the data is no longer necessary (goods or service delivered, invoice paid), it may no longer be processed on this ground. Further processing will require an other ground, for example, consent (for another purpose);

c. processing is necessary for **compliance with a legal obligation** to which the controller is subject → much processing is mandatory due to legal obligations, such as processing by tax authority, social security agency, land registry, or by commercial enterprise that must, for example, comply with employment, social security, and tax legislation. Article 6.3 stipulates that this processing must be based on Member States or Union Law, must contain the specific purpose(s), and must have relevant limitations and safeguards;

d. processing is necessary in **order to protect the vital interests** of the data subject or of another natural person → this ground must be understood as concerning life-threatening situations, where, for example, medical data must be processed to save someone's life;

e. Processing is necessary for the **performance of a task carried out in the public interest** or in the **exercise of official authority vested in the controller** → we can think of processing by various types of government agencies that provide support to those in need, or need to collect information on energy usage to develop policies on the reduction of energy consumption;

f. Processing is necessary for the purposed of the **legitimate interests pursued by the controller or by a third party, <u>except</u>** where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. Point F of the first subparagraph shall **not** apply to processing carried out by public authorities in the performance of their tasks → the f-ground is important for processing carried out by the commercial sector, including financial institutions, social networks, and search engines, and we may expect that added value service providers in the context of smart homes, smart grids, and connected cars will base the processing of data that is not necessary for the primary process (which will often be based on contract) on the f-

ground. As the economic interests of a business, including its competitive edge and innovative potential, often depend on advertising revenue and/or the sale of personal data or inferred profiles, the f-ground is a tempting basis insofar as other grounds do not apply. **However,** the f-ground requires a **balancing test**. The business interest of a company **cannot,** by default, overrule the interest of fundamental rights and freedoms of data subjects whose behavioural data are used to generate income (thus, e.g. enabling the so-called 'free services' of social networks and search engines).

The balancing test required of the controller, entails the following considerations:

- The nature and source of the legitimate interest and whether the data processing is necessary for the exercise of a fundamental right, is otherwise in the public interest, or benefits from recognition in the community concerned;
- The impact on the data subject and their reasonable expectations about what will happen to their data, as well as the nature of the data and how they are processed;
- Additional safeguards which could limit undue impact on the data subject, such as data minimisation, privacy-enhancing technologies; increased transparency, general and unconditional right to opt-out, and data portability.

The f-ground is often used to legitimize the advertising business model for free services. For instance, in the Google Spain vs Costeja case, the CJEU concluded that Google was processing personal data based on its legitimate business interests. In this case, the CJEU considered two types of legitimate interests that might overrule Costeja's interest in having a particular search result dereferenced. First, the interest of the controller, second the interest of third parties, namely the users of the search engine.

First, the Courts looks into the economic interest of Google Spain in sustaining its business model, because the right to erase and the right to object that Costeja invoked would involve costs on the side of Google. **Second,** the Court considered the legitimate interests of users of the search engine in having access to the search result that may be de-referenced.

5.5.2.6 Principles of lawful, fair, and transparent processing

Next to, and thus on top of, having a legal ground, Article 5 GDPR stipulates a set of rules under the heading of 'Principles relating to the processing of personal data'. These are rules that must be complied with:

a. Processed lawfully, fairly and in a transparent manner in relation to the data subject (**lawfulness, fairness and transparency)** → though one might think that lawfulness merely refers to Article 6, which contains the legal basis, the term 'lawfulness' also refers to the bigger picture of the rule of law, as with the requirement that infringements of the right to privacy under Article 8 ECHR must be 'in accordance with the law'. This means that a mere basis in law is **not** enough and must be understood in qualitiv terms to include respect for legitimate expectations, independent oversight, and other checks and balances to ensure that the legal basis of Article 6 is valid. Similarly, fairness refers to various balancing and proportionality tests, taking note of the relevant interests and fundamental rights that are at stake. Transparency is further detailed in Articles 13, 14 and 15 GDPR.

b. Collected for specified, explicit and legitimate purposes and not further processed in a matter that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historic research purposes or statistical purposed shall, in accordance with Article 89 (1), **not** be considered to be incompatible with the initial purposes (**purpose limitation)** → further processing for another purpose is allowed if the purpose is **not incompatible** with the initial purpose, as communicated to the data subject. To determine whether the new purpose is compatible, Article 6 (4) provides the following indications: any link between the old and the new purpose, the context of collection and the relationship between controller and subject, the nature and sensitivity of the data, the potential consequences of further processing for the data subject, and the existence of appropriate safeguards, such as encryption or pseudonymisation. In case of consent for the new purpose or a legal obligation that involves the new purpose, processing is based on the new ground and cannot be based on processing for a compatible purpose.

Secondary usage (further processing) for scientific of statistical research of achieving in the public interest is considered compatible by default. The GDPR contains an extensive exception for such processing in Article 89, with further exceptions for medical research in, for example, Article 9.2(h). Recital 33 furthermore indicates that:

'*it is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to giver their consent to certain areas of scientific research when in keeping with recognises ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose'.*

c. Adequate, relevant and limited to what is necessary in relation to the purposes, for which they are processed (**data minimisation)** → the criterion is 'adequate, relevant and limited to what is necessary'. This is a further restriction, moving towards strict proportionality and subsidiarity, thereby also relating to the requirement to pseudonymize or anonymize the data as soon as possible. This principle links consent to necessity, as observed above. It also connects with the right to request erasure if processing is irrelevant for the giving purpose;
d. Accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (**accuracy)** → here the principle of accuracy is formulated as a legal obligation of the data controller, but this connects with the rights of erasure and rectification in the case that data are inaccurate;
e. Kept in form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processes; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject (**storage limitation)** →

storage limitation basically requires that controllers engage in lifecycle management of the personal data they process, removing them, for example, when the purpose is exhausted and processing is no longer relevant. The exception for scientific research and archiving requires appropriate technical and organizational safeguards, taking into account the right and freedoms of the data subject;

f. Processed in a manner that ensures the appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures (**integrity and confidentiality)**

Article 5 (2): the controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 (**accountability)** → the accountability principle addresses the data controller as the focal point of responsibility, accountability, and liability regarding compliance with the principles that pervade the GDPR.

5.5.2.7 Valid consent

The GDPR contains a separate article on consent. Article 7 declares, under the heading of 'Conditions for Consent':

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his of here personal data → **the burden of proof**
2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes and infringement of this Regulation shall **not** be binding → **note** that consent may not be hidden in complicated wordy privacy policies, and must be 'easily accessible' as to its form (think of the user interface), 'using clear and plain language'. If consent is part

of an elaborate and incomprehensible Terms of Service that basically contains and implicit consent, such consent is **not valid.**

3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall **not** affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent → this means that if consent is given by ticking a box, it must be as easy to untick the box. If one has to explore every nook and corner of a website to figure out how to withdraw consent, the consent is **not valid.**
4. when assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contact → to better understand what this means, we can use recital 43:

'*in order to ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was* ***freely*** *given in all the circumstances of that specific situation. Consent is presumed not to be given freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance'.*

→ This seems to indicate that attempts to force consumers to choose between accessing a service and refusing consent for additional processing are unlawful and that such consent is not valid. Additional processing refers to the processing of data that is not necessary for the provision of the service, or further processing of data after the purpose has been exhausted.

Note that the legal ground must be communicated to the data subject **when the processing commences** (if data is collected from the data subjects), or within a **reasonable time**, at the

latest within one month after obtaining the data (if data has not been obtained from the data subjects). Controllers cannot require consent and – after finding that the consent is not valid – claim that the processing is based on its legitimate interest.

5.5.4.8 Special categories of data

Article 9 defines a set of data as requiring special treatment. These data are often called 'sensitive data' and are defined as: 'data revealing racial or ethic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation'.

By default, the processing of such data is prohibited. Strictly defined exceptions apply, notably based on explicit consent; specific rights and obligations in the field of employment and social; the vital interests of the data subject or of another natural person; or with regard to processing in the context of not-for-profit bodies with a political, philosophical, religious, or trade union aim; processing of personal data which are manifestly made public by the data subject; processing necessary for legal claims, substantial public interest, preventive or occupational medicine, assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services, for public health, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. On top of that, Article 10 restricts the 'Processing of personal data relating to criminal convictions and offences'.

Article 9 and 10 demonstrate that data protection is not just about the right to privacy, but also entails protection against discrimination on prohibited grounds.

Week 3 – right to private life handbook
Stage 2 of the Article 8 test

2.1 – has there been an interference with the Article 8 right?

Once it is established that the dispute concerns private or family life, home or correspondence then the Court goes on to examine the substance of the complaint under Article 8 paragraph 2, which provides that:

*There shall be **no interference** by a public authority with the exercise of this right **except** such as in **accordance with the law** and is **necessary in a democratic society** in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others.*

What constitutes an interference?
Once it is established that the dispute concerns an Article 8 right, the next stage of the test is to determine whether the measure complained of interferes with that right. Examples include stopping prisoners' correspondence, search a person's home or removing children from their parents and taking them into public care.

What does the applicant have to establish?
It is for the applicant to establish the fact of interference. Where the applicant **cannot** establish the certainty of the material damage which could constitute the interference, it will be sufficient if he can demonstrate a **likelihood** that the interference has occurred.

When will the existence of secret surveillance legislation be sufficient to interfere with private life?
As the authorities intend, many of the subjects of secret surveillance are oblivious to the interference. Other may suspect it, but lack sufficient proof. The applicant's difficulty in proving that this communication has been intercepted may lead him/her to claim that the mere existence of the legislation interferes with his private life and correspondence under Article 8. This claim will only succeed in certain circumstances.

In **Klass vs Germany** the Court held that an individual may claim to be victim of a violation occasioned by the mere existence of secret measures or of legislation permitting secret measures, without having to allege that such measures were in fact applied to him. However, it made it clear that this will occur under certain conditions only. The **relevant** conditions are to be determined in each case according to the Convention right alleged to have been infringed, the secret character of the measures objected to, and the connection between the applicant and those measures.

In **Malone vs the UK** there was a dispute before the court as to whether the applicant's telephone had in fact been bugged. The Court concluded that the existence of laws and practices which permit and establish a system for effecting secret surveillance of communications alone amounted to an interference with the exercise of the applicant's right under Article 8 'apart from any measures actually taken against him'.

If there has been an interference with an Article 8 right:

- *is it in accordance with the law?*
- *Does it pursue a legitimate aim?*
- *Is it necessary in a democratic society?*

Is the interference in accordance with the law?

A measure which constitutes an interference with an Article 8 right will only be compatible with that provision where it is in accordance with law. If the measure complained of does not fulfil this legal requirement it will violate Article 8 and the case will end there.

What is the meaning of 'in accordance with law'?

In order to be 'in accordance with law' the interference complained of must have a **legal basis** and the law in question bust be **sufficiently precise and contain a measure of protection against arbitrariness** by public authorities.

The interference must have a legal basis

Measures will be problematic in this regard where they are not specifically authorised by statute and are regulated instead by administrative practice, or other non-binding guidelines. An administrative practice, however well adhered to, thus does not provide the guarantee required by 'law'.

In **Malone vs UK** the Court considered whether the power to intercept telephone conversations had a legal basis. At the time, telephone-tapping was regulated by administrative practice, the details of which were not published, and without specific statutory authorisation. The Court said that there was **not a sufficient clarity about the scope or the manner in which the discretion of the authorities to listen secretly to telephone conversation was exercised**: because this was an administrative practice, it could be changed at any time and this constituted a violation of Article 8.

In **Khan vs the UK** the Court held that the use of a convert listening device by the UK authorities was not in accordance with law within the meaning of Article 8 because there was **no statutory system to regulate the use of such devices**, which was governed by Home Office Guidelines which were neither legally binding nor directly publicly accessible.

The foreseeability requirement

In order to satisfy Article 8's legality requirement, the quality of the law in question must be such that it is **accessible to the persons concerned,** and **formulated with sufficient precision to enable them,** if need be with appropriate advice, to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail. This is known as the **foreseeability requirement** and it means that a law which confers a discretion is not in itself inconsistent with Article 8 as long as the scope of discretion and the manner of its exercise are indicated with sufficient clarity to give the individual adequate protection against arbitrary interference.

Telephone tapping

In two cases against France – the **Kruslin case** and the **Huvig case** – it fell to the Court to consider whether French law regulating telephone-tapping by the police was in conformity with the foreseeability requirement of Article 8 paragraph 2. It held that:

'*tapping and other forms of interception of telephone conversations represent a serious interference with private life and correspondence and must accordingly be based on a 'law' that is particularly precise. It is essential to have clear, detailed rules on the subject, especially as the technology available for use is continually becoming more sophisticated'.*

In that respect, the Court was of the opinion that French law (written and unwritten) did not

'indicate with reasonable clarity the scope and manner of exercise of the relevant discretion conferred on the public authorities. This was truer still at the material time, so that Mr Kruslin did not enjoy the minimum degree of protection to which citizens are entitled under the rule of law in a democratic society'.

In **Rotaru vs Romania** the applicant complained that the Romanian Intelligence Service (RIS) held and used a file containing personal information on him, some of which he claimed was false and defamatory. The core issue was whether the law which permitted this interference was accessible to the applicant and foreseeable as to its results. The Court first noted that **the risks of arbitrariness are particularly great where a power of the executive is exercised in secret.** The question was whether domestic law laid down with sufficient precision the circumstances in which the RIS could store and make use of information relating to the applicant's private life. Noting that the relevant law provides that information affecting national security may be gathered, recorded and archived in secret files, the Court observed that no provision of domestic law lays down any limits on the exercise of those powers. For instance, it observed that the relevant domestic law did **not** set out any of the following:

- The kind of information that may be recorded;
- The categories of people against whom surveillance measures such as gathering and keeping information may be taken;
- The circumstances in which much measures may be taken; and
- The procedure to be followed.

Nor did it place any limits on the age of information held or the length of time for which it may be kept. Moreover, in relation to the safeguards which were necessary to protect against arbitrary use of power to gather and archive information, the Court noted that Romanian law did not provide any supervision procedure. Overall, then, it was found not to indicate with reasonable clarity the scope and manner of exercise of the relevant discretion conferred on the public authorities and the holding and use by the RIS of information on the applicant's private life was thus **not** 'in accordance with the law', and in violation of Article 8.

Does the interference pursue a legitimate aim?
Once an interference is found to be in accordance with law, the Court will proceed the question whether it pursues a legitimate aim under Article 8 paragraph 2. Article 8 paragraph 2 contains a list of the aims upon which the state can seek to rely in this regard. For example, the State may argue that:

- The collection and storage of information about individuals is 'in the interests of national security';
- Intercepting prisoners' correspondence seeks to prevent 'disorder and crime';
- Removing children from an abusive home or denying one party custody or contact aims to protect 'health or morals' or the 'rights and freedoms of others'.

It falls on the respondent State to identify the objective or objectives of the interference, and the fact that the grounds for permissible interference are so wide – in the interests of national security for example – means that the State can usually make a plausible case in support of the interference.

Is the interference necessary in a democratic society?

What is the meaning of 'necessary'?
It is clearly **not** sufficient that the State had 'some' reason for taking the measures that created the interference as the interference must be 'necessary', the Court explained in **Handyside vs the UK:**
'it is not synonymous with 'indispensable', neither has it the flexibility of such expressions as 'admissible', 'ordinary', 'reasonable', or 'desirable'.
The Court elaborated further in **Olssen vs Sweden:**
*'the notion if necessity implies that an interference corresponds to a **pressing social need** and, in particular, that it is proportionate to the legitimate aim pursued.*

The character of a democratic society?
In **Dudgeon vs the UK** the Court spoke of tolerance and broadmindedness as two of the hallmarks of a democratic society. What is necessary in a democratic society for the purposes of Article 8 is determined by reference to the balance achieved between the rights of the individual and the public interest, through the application of the principle of **proportionality.**

What is the principle of proportionality?
Overall, the principle of proportionality recognises that human rights are **not absolute** and that the exercise of an individual's rights must always be checked by the broader public interest. This principle is one way in which this balance is achieved and its use throughout the Court's application of the Convention is now widespread. The Court has frequently reminded that:
Inherent in the whole of the Convention is a search for a fair balance between the demands of the general interest of the community and the requirements of the protection of the individuals' fundamental rights.

How is the principle of proportionality applied to Article 8?
In carrying out its review of whether domestic decisions are compatible with Article 8, the Court applies the proportionality test, which, as it simplest, involves **balancing the rights of the individual and the interests of the State.** The Court does not offer an appeal from the decisions of domestic, however, and it thus refrains from substituting in its opinion on the merits of any individual case. Instead, its role is to consider whether, in the light of the case

as whole, the authorities had **relevant and sufficient reasons** for taking the contentious measures.

Deciding whether the interference is proportionate to the aim which it pursues is frequently a complex process, which involves consideration of a number of factors. These include the interest to be protected from interference, the severity of the interference and the pressing social need which the State is aiming to fulfil.

In relation to the **interest to be protected from interference,** for example, the Court noted in **Dudgeon vs the UK** that the right to private enjoyment of sexual relations required 'particularly serious reasons' to justify interference with it. Some rights will thus inevitably be afforded more importance than others, making interferences with them very difficult to justify.
With regard to the **nature of the interference,** it is clear that the more far reaching and severe the interference, the stronger the reasons required to justify it.

The **pressing social need** served by interference will also require serious consideration and measures used to protect national security may be easier to justify than those seeking to protect morals, for example.

The margin of appreciation
It is clear that the Court affords to the State a margin of appreciation when deciding whether an interference with an Article 9 right is justified under paragraph 2 of that provision. The margin of appreciation afforded to competent national authorities will vary according to the circumstances, the subject matter and its background. It has already been identified that factors to be taken into account in determining the scope of the margin of appreciation in this area include the existence of common ground among the laws of Contracting States; the sensitivity of the area being considered and the variety in customs, policies and practices across Contracting States.

As a rule, then, the scope of the margin of appreciation will differ according to the context. For example, it has been held to be particularly wide in areas such as child protection. Here, the Court has recognised that there is diversity in approached to child care and state intervention into the family among Contracting States, and it takes this into account when examining such cases under the Convention by allowing States a measure of discretion when acting in this area.

Moreover, the Court has also recognised that due to their proximity to the sensitive and complex issues being determined at national level, the domestic authorities are better placed to make an assessment of the circumstance of each case and to determine the most appropriate course of action.

Week 3 – key principles of European data protection law

Article 5 of the GDPR sets out the principles governing the processing of personal data. These principle cover:

- lawfulness, fairness and transparency;

- purpose limitation;
- data minimisation;
- data accuracy;
- storage limitation;
- integrity and confidentiality.

The lawfulness, fairness and transparency of processing principles

The principle of lawfulness, fairness and transparency apply to **all** personal data processing. Under the GDPR, lawfulness requires either:

- consent of the data subject;
- necessity to enter a contract;
- a legal obligation;
- necessity to protect the vital interests of the data subject or of another person;
- necessity for performing a task in the public interest;
- necessity for the legitimate interests of the controller or a third party, if they are not overridden by the interests and rights of the data subject.

Personal data should be done in a **fair manner.** The data subject must be informed of the risk to ensure that processing does not have unforeseeable negative effects. Personal data processing should also be done in a **transparent manner.** The controller must inform data subjects before processing their data, among other details, about the purpose of processing and about the identity and address of the controller. Information on processing operations must be provided in clear and plain language to allow data subjects to easily understand the rules, risks, safeguards and rights involved. Data subjects have the right to access their data wherever they are processed.

Lawfulness of processing

Lawful processing requires the **consent** of the data subject or another legitimate ground provided in the data protection legislation. Article 6(1) of the GDPR includes 5 lawful ground for processing, in addition to consent, i.e. when processing personal data is necessary for the performance of a contract, for the performance of a task carried out in the exercise of public authority, for compliance with a legal obligation, for the purpose of the legitimate interests of the controller of third parties, or if necessary to protect in the vital interests of the data subject.

Fairness of processing

The principle of fair processing governs primarily the relationship between the controller and data subject.

Controllers should notify data subjects and the general public that they will process data in a lawful and transparent matter and must be able to demonstrate the compliance of processing operations with the GDPR. Processing operations must **not** be performed in secret and data subjects should be aware of potential risks. Furthermore, controllers, so far as possible, must act in a way which promptly complies with the wishes of the data subject, especially where his or her consent forms the legal basis for the data processing.

In relation to internet services, the features of data processing systems must make it possible for data subjects to really understand what is happening with their data. In any case, the principle of fairness goes beyond transparency obligations and could also be linked to processing personal data in an ethical manner.

Transparency of processing

EU protection laws require personal data processing to be done 'in a transparent manner in relation to the data subject'. This principle establishes an obligation for the controller to take any appropriate measure in order to keep the data subjects – who may be users, customers or clients – informed about how their data are being used. Transparency may refer to the information given to the individual before the processing starts, the information that should be readily accessible to data subjects during the processing, but also to the information given to data subjects following a request of access to their own data.

Processing operations must be explained to the data subjects in an easily accessible way which ensures that they understand what will happen to their data. This means that the specific purpose of processing personal data must be known by the data subject at the time of the collection of the personal data. The transparency of processing requires that **clear and plain language** be used. It must be clear to the people concerned what the risks are, the rules, safeguards and rights regarding the processing of their personal data.

The principle of purpose limitation

The principle of purpose limitation is one of the fundamental principles of European data protection law. The principle requires that any processing of personal data must be done for a specific, well-defined purpose and only for additional purposes that are compatible with the original purpose. The processing of personal data for undefined and/or unlimited purposes is thus unlawful. The legitimacy of processing personal data will depend on the purpose of the processing, which must be explicit, specified and legitimate.

Every new purpose for processing data which is **not** compatible with the original one must have its own particular legal basis and cannot rely on the fact that the data were initially acquired or processed for another legitimate purpose.

When considering the scope and limits of a particular purpose the GDPR telly on the concept of compatibility: the use of data for compatible purposes is allowed on the grounds of the initial legal basis. Further processing of the data may **not,** therefore, be done in a way that is unexpected, inappropriate or objectionable for the data subject. To assess whether the further processing is to be considered compatible, the controller should take the following into account (among other things):

- any link between those purposes and the purposes of the intended further processing;
- the context in which the personal data have been collected, in particular concerning the reasonable expectations of data subjects based on their relationship with the controller on its further use;

- the nature of the personal data;
- the consequences of the intended further processing for data subjects; and
- the existence of appropriate safeguards in both the original and intended further processing operations. This could be done, for instance, through encryption or pseudonymisation.

The GDPR declares that 'further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' is a priori considered compatible with the initial purpose. However, appropriate safeguards such as the anonymization, encryption or pseudonymisation of the data, and restriction of access to the data, must be in place when further processing personal data. The GDPR adds that 'where the data subject has given consent or the processing is based on Union of Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard, in particular, important objectives of general public interest, the controller should be allowed to further process the personal data irrespective of the compatibility of the purposes. When undertaking further processing, the data subject should therefore be informed of the purposes, as well as his or her rights, such as the right to objects.

The data minimisation principle

Data processing must be limited to what is necessary to fulfil a legitimate purpose. The processing of personal data should only take place when the purpose of the processing cannot be reasonably fulfilled by other means. Data processing may not disproportionately interfere with the interests, rights, and freedoms at stake.

Only such data shall be processed as 'adequate, relevant and not excessive in relation to the purpose for which they are collected and/or further processed'. The categories of data chosen for processing must be necessary in order to achieve the declared overall aim of the processing operations, and a controller should strictly limit collection of data to such information as is directly relevant for the specific purpose pursued by the processing.

The data accuracy principle

The principle of **data accuracy** must be implemented by the controller in all processing operations. Inaccurate data must be erased or rectified without delay. Data may need to be checked regularly and kept up to data to secure accuracy.

A controller holding personal information shall **not** use that information without taking steps to ensure with reasonable certainty that the data are accurate and up to date.

The obligation to ensure accuracy of data must be seen in the context of the purpose of data processing.

There may also be cases where updating stored data is legally prohibited, because the purpose of storing the data is in principally to document events as a historical snap-shot.

Example: a medical record of an operation must **not** be changed, in order words 'updated', even if findings mentioned in the record later on turn out to have been wrong. In such circumstances, only additions to the remarks in the record may be made, as long as they are clearly marked as contributions made at a later stage.

On the other hand, there are situations where it is absolute necessity to update and regularly check the accuracy of data, due to the potential damage which might be caused to the data subject if data were to remain inaccurate.

Example: if somebody wants to conclude a credit contract with a banking institution, the bank will usually check the creditworthiness of the prospective customer. For this purpose, there are special databases available containing data on the credit history of private individuals. If such a database provides incorrect or outdated data about an individual, this person may suffer negative effects. Controllers of such databases must therefore make special efforts to follow the principle of accuracy.

The storage limitation principle

The principle of storage limitations means that personal data must be deleted or anonymised as soon as they are no longer needed for the purposes for which they were collected.

Article 5(1)(e) GDPR requires personal data to be 'kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data' are processed. The data must therefore be erased or anonymised when those purposes have been served. To this end, 'time limits should be established by the controller for erasure or for a periodic review' to make sure that the data are kept for no longer than is necessary.

In **S. and Marper,** the ECtHR concluded that the core principles of the relevant instruments of the Council of Europe, and the law and practice of the other Contracting Parties, required data retention to be proportionate in relation to the purpose of collection and limited in time, particularly in the police sector.

For example: in S. and Marper, the ECtHR ruled that indefinite retention of the fingerprints, cell samples and DNA profiles of the two applicants was disproportionate and unnecessary in a democratic society, considering that the criminal proceedings against both applicants has been terminated by an acquittal and a discontinuance, respectively.

The time limitation for storing personal data only applies to data kept in a form which permits identification of data subjects. Lawful storage of data which are no longer needed could, therefore, be achieved by anonymising data.

Archiving data for public interest, scientific or historical purposes, or for statistical use, may be stored for longer periods, providing such data will be used solely for the above principles.

The data security principle

The security and confidentially of personal data are key to preventing adverse effects for the data subject. Security measures can be of a technical and/or organisational nature. Pseudonymisation is a process that can protect personal data. The appropriates of security measures must be determined on a case-by-case basis and reviewed regularly.

The principle of data security requires that appropriate technical or organisational measures are implemented when processing personal data to protect the data against accidental, unauthorised or unlawful access, use, modification, loss, destruction or damage. The GDPR states that the controller and the processor should take into account 'the state of the art, the costs of implementation and te nature, scope, context and purpose of processing, as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons' when implementing such measures. Depending on the specific circumstances of each case, appropriate technical and organisational measures could include, for example, pseudonymising and encrypting personal data and/or regularly testing and evaluating the effectiveness of the measures to ensure the data processing is secure.

Pseudonymising data means **replacing** the attributes in personal data – which makes it possible to identify the data subject – with a pseudonym, and keeping those attributes separate, under technical or organisational measures. The process of pseudonymisation must **not** be confused with the process of anonymization, where all links to identifying the person are broken.

In its Opinion on the Data protection implications of the processing of Passenger Name Records, the Council of Europe provides other examples of appropriate security measures for the protection of personal data in passenger name record systems. These include holding data in a secure physical environment, limiting access control via layered logins and protecting the communication of data with strong cryptography.

In cases where a personal data breach takes place, the GDPR requires the controller to notify the competent supervisory authority of the breach with risks for rights and freedoms of individuals without undue delay. A similar communication obligation to the data subject exists when the personal data breach is likely to result in a high risk to his or her rights and freedoms. Communication of such breaches to the data subject must be in clear and plain language. If the processor becomes aware of a personal data breach, the controller must be notified immediately. In certain situations, exceptions to the notification obligation may apply. For instance, the controller is **not** required to notify the supervisory authority when 'the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons'. Nor is it necessary to notify the data subject when implemented security measures render the data unintelligible for non-authorised persons or when subsequent measures ensure that the high risk is no longer likely to materialise. If communication of a personal breach to the data subjects would involve disproportionate effort on behalf of the controller, a public communication or similar measure can ensure that 'the data subjects are informed in an equally effective manner'.

The accountability principle

Accountability requires controllers and processors to actively and continuously implement measures to promote and safeguard data protection in their processing activities. Controllers and processors are responsible for compliance of their processing operations with data protection law and their respective obligations. Controllers must be able to demonstrate compliance with data protection provisions to data subjects, the general public and supervisory authorities at any time. Processors must also comply with some obligations strictly linked to accountability (such as keeping a record of processing operations and appointing a Data Protection Officer).

The GDPR sets out that the controller is responsible for, and must be able to demonstrate compliance with, the personal data processing principles described in this chapter. Even though the accountability principle in Article 5 (2) of the GDPR is only directed towards controllers, processors are also expected to be accountable, given that they have to comply with several obligations and that they are closely connected to accountability.
While the principle of accountability in Article 5 (2) of the GDPR is not specifically directed to processors, there are provisions linked to accountability that also contain obligations for them, such as keeping a record of processing activities and appointing a Data Protection Officer for any processing activities that require one. Processors must also ensure that all measures necessary for ensuring the security of the data have been implemented. The legally binding contract between the controller and the processor must set out that the processor shall assist the controller in some of the compliance requirements, such as when carrying out a data protection impact assessment or notifying the controller of any personal data breach as soon as they become aware of it.

According to Article 29 Working Party's opinion, the essence of accountability is the controller's obligation to:

- put in place measures which would – under normal circumstances – guarantee that data protection rules are adhered to in the context of processing operations; and
- have documentation ready which demonstrates to data subjects and to supervisory authorities the measures that have been taken to achieve compliance with the data protection rules.

The principle of accountability thus requires controllers to actively demonstrate compliance and **not** merely wait for data subjects or supervisory authorities to point out shortcomings.

Chapter 4 – Rules of European data protection law
Rules on lawful processing

Personal data may be lawfully processed if they meet one of the following criteria:

- the processing is based on the consent of the data subject;
- a contractual relationship requires the processing of personal data;
- the processing is necessary for compliance with a legal obligation of the controller;
- vital interests of data subjects or of another person require the processing of their data;
- the processing is needed for the performance of a task in the public interest;

- legitimate interests of controllers or third parties are the reason for processing, but only as long as they are not overridden by the interests or the fundamental rights of the data subjects

Lawful processing of sensitive personal data is subject to a special, **stricter** regime.

Lawful grounds for processing data

Chapter II of the GDPR, entitled 'Principles', provides that all personal data processing must comply, firstly, with the principles relating to data quality set out in Article 5 GDPR. One of the principles is that personal data should be 'processed lawfully, fairly and in a transparent way'. Secondly, for data to be processed lawfully, the processing must comply with one of the lawful grounds for making data processing legitimate, listed in Article 6 for non-sensitive personal data, and in Article 9 for special categories of data (or sensitive data).

Under EU law, consent as a basis for lawful data processing is firmly established in Article 6 of the GDPR and is also explicitly referred to in Article 8 of the Charter. The characteristics of valid consent are explained in the definition of consent in Article 4, while the conditions for obtaining valid consent are detailed in Article 7 and the special rules for child's consent in relation to information society services are established in Article 8 of the GDPR.

Free consent

EU law stipulates that consent is **not** considered freely given 'if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment.' The GDPR stresses that 'when assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract'.

Free consent could also be in doubt in situations of subordination, where there is a significant economic or other imbalance between the controller securing consent and the data subject providing consent. A typical **example** of such imbalances and subordination is an employer's processing of personal data, within the context of an employment relationship. According to the Article 29 Working Party, 'employees are almost never in a position to freely give, refuse or revoke consent, given that dependency that results from the employer/employee relationship. Given the imbalance of power, employees can only give free consent in exceptional circumstances, when no consequences at all are connected to acceptance or rejection of an offer.'

This does however not mean, that consent can never be valid in circumstances where 'not consenting' would have some negative consequences. For instance, if not consenting to having a supermarket's customer card only results in not receiving a small reduction in the price of certain goods, consent could be a valid legal basis for processing the personal data of those customers who consented to having such a card. There is no subordination between company and customer and the consequences of not consenting are not serious enough to

prevent the data subject's free choice (provided that the price reduction is small enough not to affect their free choice).

However, where goods or services can only be obtained if certain personal data are disclosed to the controller or further on to third parties, the data subject's consent to disclose their data, which are not necessary for the contract, **cannot** be considered a free decision and is, therefore, not valid under data protection law. The GDPR is rather strict in forbidding the bundling of consent with the provision of goods and services.

Informed consent

The data subject must have sufficient information before exercising his or her choice. Informed consent will usually comprise a precise and easily understandable description of the subject matter requiring consent. As the Article 29 Working Party explains, consent must be based upon an appreaciation and understanding of the facts and implications of the data subject's action to consent to the processing. Therefore, 'the individual concerned must be given, in a clear and understandable manner, accurate and full information of alle relevant issues, such as the nature of the data processed, purposes of processing, the recipients of possible and the rights of the data subject'. For consent to be informed, individuals must also be aware of the consequences of not consenting to processing.

The recitals of the GDPR stipulate that informed consent means that 'the data subject should be **aware** at least of the identity of the controller and the purposes of the processing for which the personal data' processed are intended.

The quality of the information is important. Quality of information means that the information's language should be adapted to its foreseeable recipients. Information must be given without jargon, in a clear and plain language that a regular user should be able to understand. Information must also be easily available to the data subject and can be provided orally or in writing. Accessibility and visibility of the information are important elements; the information must be clearly visible and prominent.

Specific consent

For consent to be valid, it must also be specific to the processing purpose, which must be described clearly, and in unambiguous terms. The data subject must be asked again for consent if processing operations are to be added or changed in a way which could not have reasonably been foreseen when the initial consent was given and thus lead to a change of purpose. When the processing has multiple purposes, consent should be given for all of them.

Unambiguous consent

All consent must be given in an **unambiguous** way. This means that there should be **no** reasonable doubt that the data subject wanted to express his or her agreement to allow the processing of his or her data. For instance, inactivity from a data subject does not indicate unambiguous consent.

If consent is given in a written form which is part of a contract, consent for processing personal data must be individualised and in any case 'safeguards should ensure that the data subject is aware of the facts that and the extent to which consent is given'.

Consent requirements for children

The GDPR provides specific protection for children in the context of providing information society services, because 'they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data'. Therefore, under EU law, when providers of information society services process personal data of children under the age of 16 years on the basis of consent, such processing will be lawful 'only if, and to the extent that, consent is given or authorised by the holder of parental responsibility over the child'. Member States may provide for a lower age in national law, though not lower than 13 years.

The right to withdraw consent at any time

The GDPR includes a general right to withdraw consent at any time. The data subject must be informed of such a right prior to giving consent and he or she may exercise this right at his or her discretion. There should be no requirement to give reasons for withdrawal and no risk of negative consequences over and above the termination of any benefits which may have derived from the previously agreed data use. Withdrawing consent should be easy as giving it. There can be **no** free consent if the data subject is unable to withdraw his or her consent without detriment or if withdrawal is not as easy as giving consent had been.

Necessity for the performance of a contract

Under EU law, Article 6 (1) (b) GDPR provides another basis for legitimate processing, namely if it is 'necessary for the performance of a contract to which the data subject is party'. This provision also covers pre-contractual relationships. For instance, in cases where a party intends to enter into a contract, but not has yet done so, possibly because some checks remain to be completed. If one party needs to process data for this purpose, such processing is legitimate as long as it is 'necessary in order to take steps at the request of the data subject prior to entering into a contract'.

Legal duties of the controller

EU law sets out another ground for making data processing legitimate, namely if 'it is necessary for compliance with a legal obligation to which the controller is subject' (Article 6 (1)(c) GDPR). This provision refers to controllers acting in both the private and public sector; the legal obligations of public sector data controllers can also fall under Article 6 (1) (e) of the GDPR. For instance, employers must process data about their employees for social security and taxation reasons, and businesses must process data about their customers for tax purposes.

Vital interests of the data subject or those of another natural person

Under EU law, Article 6 (1) (d) GDPR provides that personal data processing is lawful if it 'is necessary in order to protect the vital interests of the data subject or of another natural person'. This legitimate ground may only be invoked for processing personal data based on the vital interests of another natural person, if such processing 'cannot be manifestly based on another legal basis'.

Public interest and exercise of official authority

Given the many possible ways of organising public affairs, Article 6 (1) (e) GDPR provides that personal data may lawfully be processed if it 'is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller'.

Legitimate interests pursued by the controller or by a third party

Under EU law, the data subject is **not** the only one with legitimate interests. Article 6 (1) (f) GDPR provides that personal data may lawfully be processed if it 'is necessary for the purposes of the legitimate interests pursued by the controller or by the third party of parties [except public authorities in the performance of their tasks] to whom the data are disclosed, except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection.

The existence of a legitimate interest must be carefully assessed in each specific case. If the legitimate interests of the controller are identified, then a balancing exercise must be conducted between those interests and the interests or fundamental rights and freedoms of the data subject. The reasonable expectations of the data subject must be considered during such an assessment to ascertain whether the interests of the controller override the interests or fundamental rights of the data subject. If the data subject's rights override the controller's legitimate interests, then the controller can take measures and implement safeguards to ensure that the impact on the data subject's right is minimised (such as pseudonymising data), and invert the 'balance' before being able to lawfully rely on this legitimate basis for processing. In its Opinion on the notion of legitimate interests of the data controller, the Article 29 Working Party underlined the crucial role of accountability and transparency, and of the data subject's rights to object to the processing of their data, or to it being accessed, modified, deleted or transferred, when balancing the legitimate interests of the controller and the interests of the data subject's fundamental rights.

In the GDPR recitals, some examples are given as to what constitutes a legitimate interest of the data controller concerned. For instance, the processing personal data is allowed without the data subject's consent when it is done for direct marketing purposes or when such processing is 'strictly necessary for the purposes of preventing fraud'.

Example: the Valsts polijas Rigas Regiona parvaldes Kartibas policijas parvalde case concerned damage to a Rigas Transport Company trolleybus caused by a passenger suddenly opening a taxi door. Rigas satiksme wanted to sue the passenger for damages. However, the police would only give the name of the passenger and refused to provide the passenger's ID number

and address, arguing that the disclosure would be unlawful under national data protection laws.

The CJEU clarified that EU data protection law includes the possibility – not an obligation – of communicating data to a third party for the purposes of the legitimate interests pursued by that party. The CJEU set out **three cumulative conditions** that must be fulfilled for personal data processing to be lawful on the '**legitimate interests'** ground:

1. the third party to whom the data are disclosed must pursue a legitimate interest. In this specific case, this means that requesting personal information to sue a person for causing property damage constitutes a legitimate interest of a third party;
2. the processing of personal data must be necessary for the purposes of the legitimate interest pursued. In this case, obtaining personal information such as the address and/or ID number is strictly necessary to identify that person;
3. the fundamental rights and freedoms of the data subject must not take precedence over the controller's or third parties' legitimate interests. The balance of interests must be done on a case-by-case basis, taking into account elements such as the severity of the infringement of the data subject's rights or even the age of the data subject in certain circumstances. However, in this specific case the CJEU did not consider the refusal of disclosure to be justified simply because the data subject was a minor.

Whenever personal data is processed under the 'legitimate interests' ground, the individual has the right to object at any time to the processing, on grounds relating to his or her particular situation, according to Article 21 (1) GDPR. The controller must stop the processing, unless it demonstrates compelling legitimate ground to continue it.

Processing special categories of data (sensitive data)

Article 9 GDPR contains a detailed regime for processing special categories of data (also called 'sensitive data'). These data reveal racial or ethnic origin, political opinions, religious or philosophical beliefs and trade union membership as well as for processing genetic and biometric data for the purposes of uniquely identifying a natural person, and for data concerning health, a person's sex life or sexual orientation. **The processing of sensitive data is prohibited in principle.**

Week 4 – circumstances by design – dark patterns in cookie consent for online news outlets

To ensure that users of online services understand what data are collected and how they are used in algorithmic decision-making, the GDPR specifies informed consent as a minimal requirement.

In practice, the regulation of data collection and privacy for online services requires, not only that the process of automation and augmentation is adequately managed, but also that the services offer a carefully designed interface. These interfaces are not a neutral conduit but can help or hinder the users in acting in their own best interest.
The term **dark patterns** have been coined to identify 'instances where designers use their knowledge of human behaviour (e.g. psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest'.

Background and related work

We are here concerned specifically with the design of consent elicitation website elements, which are sometimes called cookie consent notices. These elements are implemented as a pop-up or as a banner or panel that is part of the website. Cookie consent notices are typically implemented to demonstrate compliance with the GDPR. The GDPR refers to interaction between humans and computational systems in which human consent for the operation of the system is elicited, specifically article 32.

Dark patterns as a term was introduces by Harry Brignull who defines them as 'tricks used in websites and apps that make you things that you didn't mean to'. The concept of Dark patterns in user experience design was refined by Gray who specified 5 different types of dark patterns:

1. **Nagging:** a minor redirection of expected functionality that may persist over one or more interactions. Nagging often manifest as a repeated intrusion during normal interaction, where the user's desired task is interrupted one or more times by other tasks not directly related to the one the user is focusing on;
2. **Obstruction:** impending a task flow, making an interaction more difficult than it inherently needs to be with the intent to dissuade an action. Obstruction often manifests as a major barrier to a particular task the user may want to accomplish;
3. **Sneaking:** an attempt to hide, disguise, or delay the divulging of information that has relevance to the user. Sneaking often occurs in order to make the user perform an action they may object to if they had the knowledge
4. **Interface interference:** any manipulation of the user interface that privileges specific actions over others, thereby confusing the user or limiting discoverability of important action possibilities. Interface interference manifests as numerous individual visual and interactive deceptions;
5. **Forced action:** any situation in which users are required to perform a specific action to access (or continue to access) specific functionality. This action may manifest as a required step to complete a progress, or may appear disguised as an option that the user will greatly benefit from.

Utz have conducted a field study of consent notices on a live (e-commerce) website to identify how the design of notice influence does the user decision to accept the website cookies. They studied 80.000 unique users. Specifically, they looked into the relative position of the notice, use of nudging and the presence of a privacy link (that explains in detail how data is collected and used) and showed that small UI design decisions substantially impact whether and how people interact with cookie consent notices. One of their experiments indicated that nudging via interface interference (highlighting Accept buttons in a binary choice with decline) and pre-selected choices for different uses of cookies has a strong impact on the ratio of users who give content to accept the third-party cookies.

Nouwens also ran a user analyses, on 40 participants on the effect the consent notice design has on whether consent is given. They found that there was an approximate 22% increase in acceptance when the opt-out option was 'hidden' behind the initial notice (at least two clicks are needed to opt out)

It should be noted that the ambiguity of the term 'informed consent' is in itself an issue. Nouwens makes considerable effort to identify the features of the legal understanding of consent in the laws of the European Union. They work with the definition of Article 4 (11) GDPR: 'any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her'. Utz conclude that opt-out consent banners are unlikely to produce intentional or meaningful expression of consent.

Existence and examples

- **Nagging:** the website tried to push the visitor to change their mind by displaying a 'are you sure' notice;
- **Obstruction:** the most common obstruction pattern is hiding the option to deny consent in a page separate from the consent notice and behind obfuscating text such as 'find out more'. The user can only, for example, opt out of consent when making browser adjustments which they can find about after following the 'use of cookies' link;
- **Sneaking:** this dark pattern can be most clearly seen in widgets that state 'by continuing to use our website, we assume you accept our policies';
- **Interface interference:** the clearest example of interference is having the option to deny consent hidden by design. For example, when you have to click on the 'learn more' button to the actual notice that allows the user to deny consent;
- **Forced action:** forced actions patterns are observed in consents implemented as screen popups that block the users from accessing the website and require the users to click on its before they can continue. Geo blocking can also be seen as form of forced action.

How should regulators state their requirements?
One concrete requirement would be that the consent acceptance and denial use the same widget on the same level – the option for acceptance should be next to the option for rejection in the same design.

Ideally, we need to further define the concept and types of dark patterns, perhaps for a specific context at the time, in such a way that:

- The features that characterise a dark pattern are clearly identifiable,
- The characterising features are easily computer-detectable.

Having such a dark pattern definition would enable regulators to automatically flag violators, which in turn, we expect, would increase compliance. To contribute towards this goal, we highlight the patterns of misdirection we have identified, which can be seen as a refinement of the interface interference, obstruction and forced action patterns of Gray:

- **Does not count:** Matte have indicated that although consent has not been given (yet of has been denied) data is collected anyway. This can be seen as a dark pattern and it is specific since it can only be computationally detected (by following what the browser does on the back-end);
- **No choice:** all the included links and buttons lead to a page that either instructs to further pages that detail adjustments of browser settings, direct the user to contact third-party services or just 'explains' cookies and purposes;

- **Multiple choice panels:** the user should be asked for consent in only one notice panel. This may appear obvious, but we have detected examples where the consent can be given in two panel, but the denial option (if offered at all) is given in only one, smaller, panel;
- **Choice cascade:** the denial of consent is only reached by following a number of links or buttons that offer more information. One example is: after following the learn more link the panel with the 'Decline' option appears;
- **Widget inequality:** widget inequality is when the execution of giving consent is made easy (bright button clearly labelled with a positive word), the denial of consent option, even if given in the same panel as the acceptance is given a different design;
- **Unlabelled sliders:** the consent notices use sliders to allow users to consent or not to individual services, but it is not labelled which side of the panel is accept/on/active or which is reject/off/inactive;
- **Unmarked X:** when the panel has (usually) a top right 'x' widget, but the panel text does nog explain whether clicking this 'x' counts as consent or denial of consent;
- **No antonyms:** lastly, the use of clear words such as 'I agree', 'I consent', 'yes' to label consent option and not using their antonyms to label the denial of consent option.

Lastly, we put forward a design suggestion for a cookie consent notice that we believe fulfils the requirements of the GDPR and does not have dark patterns. As the GDPR requires the user must give an informed consent for the different types of data collected. This requirement is fulfilled by giving the user the opportunity to select which data categories they want to consent to, if any. Only 'Necessary' is pre-selected, which the GDPR allows as these are cookies that are needed for the website to function. Further, all cookie options are displayed at the first page: the user can both accept or deny all cookies with one click. All the buttons have the same size and colour as **not** to indicate to the user that one alternative is more correct than others. The wordings on the different buttons are made as clear as possible not to confuse the user about their purpose.

It is important that the cookie consent notice does not force the user to make a choice. The cookie consent notice must therefore be placed accordingly. Lastly, it is important that the user know how to change the settings later on. The cookie consent notice should include this type of information and the website should make this information accessible in their website.

Week 5 – privacy by design

Big data analytics

The term **big data analytics** refers to the whole data management lifecycle of collecting, organizing and analysing data to discover patterns, to infer situations or states, to predict and to understand behaviours. Its value chain includes a number of phases that can be summarised as follows:

- **Data acquisition/collection:** the process of gathering, filtering and cleaning data before it is put in a data repository or any other storage solution on which data analysis can be carried out. Examples of potential sources are social networks, mobile apps, wearable devices, smart grids, online retail services, public registers etc. As the main purpose is to maximise the amount of available data (so as to appropriately feed the

analyses), the process is usually based on fast and massive data collection, thus, assuming high-volume, high-velocity, high-variety, and high-veracity but low-value data;
- **Data analyses:** the process concerned with making the 'raw' collected data amenable for decision-making as well as domain specific usage. The key challenge of data analysis is to find what is really useful. A critical element in that respect is to combine data from different sources in order to derive information that cannot be found otherwise;
- **Data curation:** the active management of data over its lifecycle to ensure it meets the necessary quality requirements for effective usage. It includes functions like content creation, selection, classification, transformation, validation and preservation;
- **Data storage:** storing and managing data in a scalable way satisfying the needs of applications/analytics that require access to the data. Cloud storage is the trend but in many cases distributed storage solutions would be best options (e.g. for streaming data);
- **Data usage:** covers the use of the data by interested parties and is very much dependent on the data processing scenario. For example, the results from an analysis on trends in mobile apps usage could be available for the general public or restricted to a mobile service provider who commissioned the study. Therefore, users of big data may vary from a single organisation to a wide range of parties, such as banks, retailers, advertising networks, public authorities, etc.

Big data analytics is already happening today. Many activities of our everyday life pertain analytics, which we fuel in return during some of our everyday transactions. Some usual examples include:
- In hospital, patients' vital signs can be compared against historical data to identify patterns and provide valuable information for early detection and treatment of diseases;
- Mobile fitness apps collect data on the way we walk, sleeping patterns or other streams of data, which if combined with health data, can help healthcare providers offer better wellness programs.

Privacy by design in big data

Finding the appropriate mechanisms to implement privacy principles in the big data environment is the most, effective way to prevent a clash between privacy and big data which would have no winners. To this end, the concept of privacy and data protection by design is fundamental, as a mechanism to address the privacy risks from the very beginning of the processing and apply the necessary privacy preserving solutions in the different stages of the big data value chain. In this way, privacy by design can be a tool for empowering the individuals in the big data era, as well as supporting the data controllers' liability and trust.

Privacy by design: concept and design strategies

Privacy by design is neither a collection of mere general principles nor can it be reduced to the implementation of privacy enhancing technologies. In fact, it is a process involving various technological and organizational components, which implement privacy and data protection principles.

In its 2014 reports, ENISA explored the concept of privacy by design following an engineering approach. Among other things, the report, using relevant work in the field, presented eight privacies by design strategies, both data oriented and process oriented, aimed at preserving certain privacy goals:

- **Minimize:** the amount of personal data should be restricted to the minimal amount possible (data minimization);
- **Hide:** personal data and their interrelations should be hidden from plain view;
- **Separate:** personal data should be processed in a distributed fashion, in separate compartments whenever possible;
- **Aggregate:** personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful'
- **Inform:** data subjects should be adequately informed whenever processed (transparency);
- **Control:** data subjects should be provided agency over the processing of their personal data;
- **Enforce:** a privacy policy compatible with legal requirements should be in place and should be enforced;
- **Demonstrate:** data controllers must be able to demonstrate compliance with privacy policy into force and any applicable legal requirements.

Moreover, following the privacy by design strategies, the report focused on a number of privacy enhancing technologies that can be used for implementing strategies. Such technologies include authentication, attribute based credentials, secure private communications, communications anonymity and pseudonymity, privacy in databases, statistical disclosure control, privacy preserving data mining, private information retrieval, storage privacy, privacy preserving computations, transparency enhancing computations and intervenability enhancing technologies. The overall scope of the work was to bridge the gap between the legal framework and the available technologies, so as to allow for proper implementation of relevant solutions.

Privacy by design in the era of big data

Since privacy by design is about building privacy features at the very beginning of the processing, it can be a very useful concept and process for big data, allowing for early implementation of relevant controls that are protecting the individuals' personal data by default. Still, in the case of big data, due to the size and diversity of data being processed, even in near real time, serval challenges are introduced.

First of all, in order to identify trends, detect patterns and reach valuable findings, the available data sets in big data should be as rich as possible. On the other hand, data minimization and limits in data retention seem to be an integral part of a privacy by design approach. Therefore, it can be argued that data minimization contravenes big data where large volumes of data are collected and stored before being used. It is, thus, a major challenge in big data to minimize data collection, allowing at the same time for a useful rich content that can be used for analytics.

Moreover, the merging of information from different sources is also an essential part of big data analytics, going at first sight against the idea of distributed processing of personal data.

So, is the privacy by design concept viable in the age of big data? Despite some arguing that this is not totally possible, our approach in this report is to actually revert the question: can big data processing adopt to and benefit from the privacy by design approach and how can this be done in practice?

If privacy is seen as a core value of big data, privacy by design can be a powerful tool. The **minimize** strategy would then provide for better and more useful data collection. The **hide, aggregate** and **separate** strategies could allow for using personal data in analytics without affecting the individuals' private sphere. The **inform strategy** would support better mechanisms for users' information and transparency and the **control** strategy would support new practical ways for expressing consent and privacy preferences. Last, the **enforce** and **demonstrate** strategies would support the data controllers in applying their privacy policies, in line with the principle of accountability.

Design strategies in the big data analytics value chain

Data acquisition/collection

Minimize: one very important privacy principle directly related to the big data collection phase is that of data minimization. Each data controller who is collecting the data needs to precisely define what personal data are actually needed (and what is not needed) for the purpose of the processing, including also the relevant data retention periods. Specific processes should be in place to exclude unnecessary personal data from collection/transfer, reduce data fields and provide for automated deletion mechanisms.

Aggregate: in certain cases, such as in statistical analysis from distributed sources, the personal data might not even need to be collected in the first place and the collection of anonymised information might be sufficient. Local anonymization is the most prominent solution, which could allow the individual (or a controller processing data for the individual) to remove all personal information before releasing the data for analytics.

Hide: in many cases information about the individual may be collected without him/her even being aware (e.g. relating to his/her web searches and overall online behaviour). Privacy enhancing technologies that can support internet and mobile users' privacy are available today, including anti-tracking, encryption, identity, masking and secure file sharing tools.

Notice: the individuals need to be adequately informed about the collection of their personal data for big data analytics. To this end, appropriate information notices and other transparency mechanisms should be in place.

Control: again, the collection phase is the phase where the consent of the user needs to be obtained (if this is the legal basis for the processing). Practical and usable implementations of opt in mechanisms are crucial to this regard. Moreover, opt out tools should be offered to the individuals at any point of the processing.

Data analysis and curation

Aggregate: one of the most prominent techniques in the context of big data analysis is that of anonymization. Different privacy models and anonymization methods are in place to preserve data inference, for instance, in statistical disclosure control and privacy preserving data mining techniques, including association rule mining, classification and clustering. K-anonymity and differential privacy are the two main families of privacy models with different types of implementations.

Hide: another very important technique in privacy preserving analysis is encryption, especially in the context of performing searches and other computations over encrypted data so as to protect individuals' privacy.

Data storage

Hide: security measures such as granular access control and authentication are essential for protecting personal data in databases. Technologies such as Attribute Based Access Control can be much more scalable in big data, offering fine grained access control policies. Encryption is also core in protecting data at rest.

Separate: privacy preserving analytics in distributed systems are also important for the protection of personal data as they provide computations across different databases without the need for central warehouses. Access control measures and encryption techniques can again support this type of solutions.

Data usage

Aggregate: privacy preserving data publishing and retrieval are usually based on anonymization in order to prevent interference of personal data. Issues related to data provenance in the course of decision = making (bases on big data) is another topic of interest, especially regarding the credibility and the level of aggregation of mega data (so as to avoid identification of individuals)

On top of the design strategies and controls mentioned above, it is important to point out that the data controllers and processors must first of all take into account the underlying legal obligations, in particular relating to the privacy principles and the legal basis of the processing. To this end, the privacy by design enforce and demonstrate are applicable to **all** phases of the big data value chain. Automated policy definition, enforcement and compliance tools can be useful in this task, supporting liability and accountability.

Attack models and disclosure risk

In the context of anonymization, privacy can be compromised by means of two types of disclosure: identity disclosure and attribute disclosure. Most attacks and privacy paradigms can be categorized as focusing on one or the other type.

- **Identity disclosure:** an intruder is able to link the data in a published data set with a particular individual (also known as entity disclosure)
- **Attribute disclosure:** intruders improve their knowledge on the value of an attribute of an individual. Attribute disclosure can also be considered in the case of an intruder who finds out that an individual's data are included in a database.

Although it is more common that identity disclosure implies attribute disclosure, identity disclosure and attribute disclosure are two different notions. Identity disclosure does not require attribute disclosure to take place when all the available information is used in the linkage process. Attribute disclosure can be observed without identity disclosure when individuals share the same value on a given attribute.

Annex 1 – privacy and big data in smart cities: an example

Big data analytics is already happening today and in many sectors of our everyday life. In order to demonstrate the big data analytics (and the relevant privacy considerations) in practice, in this section we focus particularly on the area of smart cities, defining and further analysing three different analytics scenarios and their aggregation in the big data pool. The selection of smart cities as an example was twofold:

a. First, it is a fast-developing area of analytics that involves several cases of personal data processing, and;
b. Second, it concerns and affects a great (and increasing) number of diverse population in Europa.

Introduction to smart cities

According to the European Commission 'a smart city is a place where the traditional networks and services are made more efficient with the use of digital and telecommunication technologies, for the benefit of its inhabitants and businesses. It means smarter urban transport networks, upgraded water supply and waste disposal facilities, and more efficient ways to light and heat buildings. And it also encompasses a more interactive and responsive city administration, safe public spaces and meeting the needs of an ageing population'.

Smart city uses cases

Smart parking

Looking for an available parking space is a daily routine for drivers in big cities, which not only takes up a lot of their time but also has an impact on multiple aspects of the cities. In particular, the increased emission of gases deteriorates atmosphere's pollution and grows traffic congestion, further impacting the effectiveness of surface transportation and fuel consumption. To address this problem, city administration is deploying LoT solutions which, based on low cos sensors, can infer if a parking position is occupied.

Through real-time data collection and communication, the central system processes the sensors' measurements and, using advanced analytics, it combines occupancy information with the drivers' location sent via a smartphone application. In this way, the drivers are directed towards the parking spot that best matches their needs.

In such a scenario, many stakeholders may be involved, for example the city administration, the service provider, the telecommunication provider, the bank, an aggregator/mediator and the user (driver). One stakeholder might have more than one role, e.g. the telecommunication provider might also be the aggregator/mediator, the city administration might also be the service provider etc.

In order for the smart parking application to be effective, efficient and sustainable, the users (drivers) must disclose either prior or during the provision of a service, a number of personal data. First of all, for the real-time identification of nearby parking slots, the location of the user must be sent (via the smartphone) to the service provider. Moreover, the user might wish to declare certain preference, e.g. maximum acceptable distance of the parking slot, time availability of slot, cost of parking per hour, need for electrical outlet in case of electrical car etc. the combination and/or further analyses of these data could also provide additional information, e.g. about the user's everyday habits or contacts (if he/she often parks at a particular place at a certain time of the day).

Smart metering (smart grid)

Smart grids are energy networks that can automatically monitor energy flows and adjust to changes in energy supply and demand accordingly. Through the flow monitoring they can facilitate a smoother introduction of renewable sources of energy by accumulating their contribution to the grid, offering more stability to the electrical network and allowing grid operators to balance their networks. When combined with smart metering systems, smart grids can also provide information on real-time consumption.

Smart metering systems are connected via a telecommunication network to the metering provider, which processes the data and calculates the total cost of consumption. More advances meters can be considered as gateways, in which a number of sensors are connected (e.g. smart plugs to monitor individual devices, smart actuators to control devices, temperature and weather sensors to compensate for fluctuations in the environment, inductive sensors to the home cabinets to monitor individual electrical lines within the house, etc.).

The stakeholders involved in such a scenario include the energy supplier, the smart meter/metering service provider, the telecommunication provider, the billing provider and the user. Again, one stakeholder may have multiple roles, e.g. the energy supplier may also be offering the smart metering system.

Smart meters are based on the collection, transmission and further use of energy consumption data at a detail level much higher than that of a 'traditional' meter. These data also include user's personal data that can be inferred from the energy consumptions patterns. In particular, by analysing the energy consumption of specific appliances/devices, the user's

presence and behaviour in the house can be guessed. Moreover, when the smart meter is connected to other sensors, a detailed behavioural profile of the user could also be built (e.g. what time he/she goes to sleep or wakes up, what type of TV movies he/she is watching etc.). this potentially intrusive characteristics of smart meters have raised several privacy concerns.

Citizen platform

City administrations are promoting new ways to communicate, engage and interact with their citizens in order to be aware, in real time, of their needs and concerns, and to timely respond and deliver tailored services. One prominent way among smart city infrastructures, to facilitate this interaction, is through mobile crowdsourcing applications (hereinafter citizen platforms).

The scope of these types of applications is threefold:

a. Empower citizens to report problems that they encounter and request immediate reaction from the city officials (e.g. traffic, waste collection, water outage, emergencies, environmental violations, noise, etc.);
b. Provide personalized recommendations through city guides, entertainment and shopping suggestions based on citizens' preferences and;
c. E-participation services, allowing citizens to give their opinions regarding aspects of the city's life.

Citizen platforms usually support interoperability with social media providers, enabling users to tag, comment and share relevant information and content.

The identified stakeholders in this scenario are the citizens (users of the application), the city authorities, the telecommunication providers, the application providers and the social media providers.

Privacy considerations

Each of the above-presented scenarios have their own privacy risks that need to be appropriately managed when defining the conditions for the processing of personal data. These risks, however, can grow exponentially if data from the different use cases are correlated in the context of advanced analytics. Such correlation could be very interesting, so as to identify different behavioural patterns and graphs of citizens' movements and activities in the city (conducted for different types of purposes, e.g. demographics, provision of new services, advertising, etc.). However, if not adequately safeguarded, this type of analysis can also lead to very detailed insight in the individuals' life, including everyday movements, habits, social contacts, leisure, etc. this can be a quite typical data analytics scenario, which is not so far from being reality.

To this end, the first issue which immediately emerges in all described uses cases is **control**: who is responsible for what. The diversity and complex interrelation of the different data controllers and processors in each of the presented scenarios can create confusion and vagueness. Without a proper data protection responsibility allocation, each actor will be tempted to utilize the data according to its own economic interest, overriding users' preferences.

Another risk coming from dispersion of the roles is the potential **duplication of data.** Again, if responsibilities are not properly defined before the service is delivered to users, each actor will be tempted to store the data on its systems for its own purpose (mainly optimization and engineering), and very likely with different level of security attached to each data reservoir, depending on the economic

resources and data protection culture. It is worth reminding that in an interconnected system, the level of security is set by the weakest link, with consequences on the probability of occurrence of data breaches.

Also, and still connected to the dispersion of roles, specific risks emerge with the effectiveness of the privacy policies adopted by each actor with regard to the processing carried out by another actor. **For example,** who would be responsible for informing the data subjects and at what sage of the processing? Problems may arise in case of controversies since the stakeholders may operate under different jurisdictions.

In a scenario where many controllers are involved and personal data are collected in different ways, trust (or lack of trust) is an important factor. The richer the user profile, the higher the temptation for the operators to target a user with unsolicited advertising or to engineer a pricing structure capable to extract as much surplus from the user as possible. This practice is known as **price differentiation** and the border between differentiation and discrimination may sometimes by very thin. The possibility of data inference and re-identification cannot be excluded unless appropriate technical safeguards are in place.

Risk mitigation strategies

Following the aforementioned description, we provide a number of risk mitigation strategies that could be applied in the context of each of the smart cities big data analytics scenarios.

Transparency and awareness

Anticipating the rationale of the new GDPR, the different stakeholders may agree between themselves, and in compliance with the law, on the allocation of data protection responsibilities and on designating which of them plays the role of single point of contract for privacy related issues, as well as for any technical assistance the user may need. This will facilitate enormously the exercise of users' right and increase the effectiveness of the information. The creation of a web resource where the user can retrieve his/her own data may also be a valuable tool in this sense. **For example** in the case of citizen's platform, the application itself could offer the users' access to their posts and the uploaded images or videos, allowing also to correct and/or delete them permanently.

User control

User control needs to be reinforced, enabling users to effectively select between a number of choices (after having been informed on their impact on the user experience). Tools such as privacy preferences and personal data stores can be very interesting in this regard, proving also for smoothness and ease of use (no need to 'exist' or disrupt the service).

Data minimization

Also, in order to strengthen the data minimization principles as much as possible, the information handed over between the parties should be engineered in order to be Boolean rather than fully analytical. This is part of the data collection process and should be explicitly addressed in each different application. **For example,** in the smart parking application, instead of disclosing the entire amount of money available on the account for the payment, the binary variable Yes/No could be used.

Access control and encryption

In order to avoid crime related risks, access to data should be based on the respect of the principle of separation of duties, where each actor is enabled to have access to the data relating to the portion of

the service that he provides, on a strict need-to-know basis. **For example,** in the smart metering scenario, holding much of the data in the meter in an encrypted format, allowing the provider to query them when needed for the provision of the service, can be a strong privacy guarantee. Other technical and organizational safeguards to minimize any risk of data misuse include log management, aggregation of data whenever individual level data are not required, audit and policy enforcement.

Retention, deletion and anonymization

In each of the presented use cases, the data should be kept only for the period that are absolutely necessary for the processing. Also, upon a data subject's specific request, and if no other legitimate interests or legally binding constraints exist, personal data should be deleted. If data need to be stored beyond the data retention periods, appropriate anonymization methods need to be applied.

Privacy Impact Assessment

Finally, a practical tool in this context, and more generally in all big data applications where many stakeholders are involved, is the use of Privacy Impact Assessment (PIA) prior to the deployment of the service. In this specific context of smart grids, the European Commission has developed a PIA Template which can be usefully applied by network and smart grid system op