CHAPTER 6

THE BIG DATA WORLD: BENEFITS, THREATS AND ETHICAL CHALLENGES

Marina Da Bormida

ABSTRACT

Advances in Big Data, artificial Intelligence and data-driven innovation bring enormous benefits for the overall society and for different sectors. By contrast, their misuse can lead to data workflows bypassing the intent of privacy and data protection law, as well as of ethical mandates. It may be referred to as the 'creep factor' of Big Data, and needs to be tackled right away, especially considering that we are moving towards the 'datafication' of society, where devices to capture, collect, store and process data are becoming ever-cheaper and faster, whilst the computational power is continuously increasing. If using Big Data in truly anonymisable ways, within an ethically sound and societally focussed framework, is capable of acting as an enabler of sustainable development, using Big Data outside such a framework poses a number of threats, potential hurdles and multiple ethical challenges. Some examples are the impact on privacy caused by new surveillance tools and data gathering techniques, including also group privacy, high-tech profiling, automated decision making and discriminatory practices. In our society, everything can be given a score and critical life changing opportunities are increasingly determined by such scoring systems, often obtained through secret predictive algorithms applied to data to determine who has value. It is therefore essential to guarantee the fairness and accurateness of such scoring systems and that the decisions

Ethical Issues in Covert, Security and Surveillance Research
Advances in Research Ethics and Integrity, Volume 8, 71–91

ISSN: 2398-6018/doi:10.1108/S2398-601820210000008007

relying upon them are realised in a legal and ethical manner, avoiding the risk of stigmatisation capable of affecting individuals' opportunities. Likewise, it is necessary to prevent the so-called 'social cooling'. This represents the long-term negative side effects of the data-driven innovation, in particular of such scoring systems and of the reputation economy. It is reflected in terms, for instance, of self-censorship, risk-aversion and lack of exercise of free speech generated by increasingly intrusive Big Data practices lacking an ethical foundation. Another key ethics dimension pertains to human-data interaction in Internet of Things (IoT) environments, which is increasing the volume of data collected, the speed of the process and the variety of data sources. It is urgent to further investigate aspects like the 'ownership' of data and other hurdles, especially considering that the regulatory landscape is developing at a much slower pace than IoT and the evolution of Big Data technologies. These are only some examples of the issues and consequences that Big Data raise, which require adequate measures in response to the 'data trust deficit', moving not towards the prohibition of the collection of data but rather towards the identification and prohibition of their misuse and unfair behaviours and treatments, once government and companies have such data. At the same time, the debate should further investigate 'data altruism', deepening how the increasing amounts of data in our society can be concretely used for public good and the best implementation modalities.

Keywords: Big Data; artificial intelligence; data analytics; ethics challenges; individuals' control over personal data; dataveillance

THE ERA OF BIG DATA AND THE 'DATAFICATION' OF SOCIETY

We live in the era of Big Data, where governments, organisations and marketers know, or can deduce, an increasing number of data items about aspects of our lives that in previous eras we could assume were reasonably private (e.g. our race, ethnicity, religion, politics, sexuality, interests, hobbies, health information, income, credit rating and history, travel history and plans, spending habits, decision-making capabilities and biases and much else). Devices to capture, collect, store and process data are becoming ever-cheaper and faster, whilst the computational power to handle these data is continuously increasing. Digital technologies have made possible the 'datafication' of society, affecting all sectors and everyone's daily life. The growing importance of data for the economy and society is unquestionable and more is to come.[1]

But what does 'Big Data' mean? Though frequently used, the term has no agreed definition. It is usually associated with complex and large datasets on which special tools and methods are used to perform operations to derive meaningful information and support better decision making. However, the Big Data concept is not just about the quantity of data available, but also encompasses new ways of analysing existing data and generating new knowledge. In public discourse, the term tends to refer to the increasing ubiquity of data, the size of datasets, the growth of digital data and other new or alternative data sources. From a more specifically technical perspective, Big Data has five essential features:

- *Volume*: the size of the data, notably the quantity generated and stored. The volume of data determines its value and potential insight. In order to have Big Data, the volume has to be massive (Terabytes and Petabytes or more).[2]
- *Variety*: the type and nature of the data, as well as the way of structuring it. Big Data may draw from text, images, audio, video (and data fusion can complete missing pieces) and can be structured, semi-structured or unstructured. Data can be obtained from many different sources, whose importance varies depending on the nature of the analysis: from social networks, to in-house devices, to smartphone GPS technology. Big Data can also have many layers and be in different formats.
- *Velocity*: the time needed to generate and process information. Data have to flow quickly and in as close to real-time as possible because, certainly in a business context, high speed can deliver a competitive advantage.
- *Veracity*: data quality and reliability; it is essential to have ways of detecting and correcting any false, incorrect or incomplete data.
- *Value*: the analysis of reliable data adds value within and across disciplines and domains. Value arises from the development of actionable information.

BIG DATA AS AN ENABLER OF GROWTH BUT HARBINGER OF ETHICAL CHALLENGES

Big Data is increasingly recognised as an enabling factor that promises to transform contemporary societies and industry. Far-reaching social changes enabled by datasets are increasingly becoming part of our daily life with benefits ranging from finance to medicine, meteorology to genomics, and biological or environmental research to statistics and business.

> Data will reshape the way we produce, consume and live. Benefits will be felt in every single aspect of our lives, ranging from more conscious energy consumption and product, material and food traceability, to healthier lives and better health-care …. Data is the lifeblood of economic development: it is the basis for many new products and services, driving productivity and resource efficiency gains across all sectors of the economy, allowing for more personalised products and services and enabling better policy making and upgrading government services …. The availability of data is essential for training artificial intelligence systems, with products and services rapidly moving from pattern recognition and insight generation to more sophisticated forecasting techniques and, thus, better decision making …. Moreover, making more data available and improving the way in which data is used is essential for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and more sustainable societies. It will for example lead to better policies to achieve the objectives of the European Green Deal. (COM, 2020b)

The exploitation of Big Data can unlock significant value in areas such as decision making, customer experience, market demand predictions, product and market development and operational efficiency. McKinsey & Company (Bailly & Manyika, 2013) report that the manufacturing industry stores more data than any other sector, with Big Data (soon to be made available through Cyber-physical Systems) expected to have an important role in the fourth industrial revolution, the so-called 'Industry 4.0' (Kagermann & Wahlster, 2013). This revolution has the potential to enhance productivity by improving supply chain management (Reichert, 2014) and creating more efficient risk management systems based on better-informed decisions. Industry 4.0 is also aimed at developing intelligent products (smart products) capable of capturing and transmitting huge amounts of data on their production and use. These data have to be gathered and analysed in real-time so as to pinpoint customers' preferences and shape future products. Data are also expected to fuel the massive uptake of transformative practices such as the use of digital twins in manufacturing.

As mentioned, Big Data also creates value in many other domains including health care, government administration and education. The application of transparency and open government policies is expected to have a positive impact on many aspects of citizens' lives. This will hopefully lead to the development of more democratic and participative societies by improved administrative efficiency, alongside perhaps more obvious uses such as better disease prevention in the health sector or self-monitoring in the education sector.

However, these positive effects must be offset against complex and multidimensional challenges. In the health care sector, an area that could benefit enormously from Big Data solutions, concerns relate, for instance, to the difficulty of respecting ethical boundaries relating to sensitive data where the volume of data may be preventing the chance to acquire the informed and specific consent required before each processing instance takes place. Another example, in the education sector, is the risk that students feel under surveillance at all times due to the constant collection and processing of their data, thus potentially leading to a reduction of their creativity and/or in higher levels of stress.

When considering Big Data, the debate needs to highlight the several potential ethical and social dimensions that arise, and explore the legal, societal and ethical issues. Here, there is a need to elaborate a societal and ethical framework for safeguarding human rights, mitigating risks and ensuring a consistent alignment between ethical values and behaviours. Such a framework should be able to enhance the confidence of citizens and businesses towards Big Data and the data economy. As acknowledged by the European Data Protection Supervisor (EDPS), 'big data comes with big responsibility and therefore appropriate data protection safeguards must be in place'.[3]

Recent ethical debate has focussed on concerns about privacy, anonymisation, encryption, surveillance and, above all, trust. The debate is increasingly moving towards artificial intelligence (AI) and autonomous technology, in line with technological advances. It is likely that as technology changes even further upcoming new types of harms may also be identified and debated.

THE CONTINUITY (OR NOT) OF DATA SCIENCE RESEARCH ETHICS WITH SOCIAL AND BEHAVIOURAL SCIENCE RESEARCH ETHICS

Given data-intensive advances, a pertinent question is whether ethical principles developed in the social and behavioural sciences using core concepts such as informed consent, risk, harm, ownership, etc. can be applied directly to data science, or whether they require augmentation with other principles specifically conceived for 'human-subjects' protection in data-intensive research activities. Traditionally, human-subjects' protection applies when data can be readily associated with the individual who bears a risk of harm in his or her everyday life. However, with Big Data there may be a substantial distance between everyday life and the uses of personal data. If technical protections are inadequate, and do not prevent the re-identification of sensitive data across distinct databases, it is challenging to predict the types of possible harms to human subjects due to the multiple, complex reasons for sharing, re-using and circulating research data.

If these difficulties are insurmountable within existing paradigms of research ethics, we will need to re-think the traditional paradigms. Here, a new framework of research ethics specific to data science could perhaps be built that could better move the 'person' to the centre of the debate. The expanding literature on privacy and other civil rights confirms that the ethical dimension of Big Data is becoming more and more central in European Union (EU) debate, and that the common goal is to seek concrete solutions that balance making the most of the value of Big Data without sacrificing fundamental human rights. Here, the Resolution on the fundamental rights implications of Big Data (2016/2225), adopted by the European Parliament, underlines that though Big Data has valuable potential for citizens, academia, the scientific community and the public and private sectors, it also entails significant risks namely with regard to the protection of fundamental rights, the right to privacy, data protection, non-discrimination and data security. The European Parliament has therefore stressed the need for regulatory compliance together with strong scientific and ethical standards, and awareness-raising initiatives, whilst recognising the importance of greater accountability, transparency, due process and legal certainty with regard to data processing by the private and public sectors.

Likewise, the European Commission (EC) recognises the importance of safeguarding European fundamental rights and values in the data strategy and its implementation (COM, 2020b), whilst in the COM (2020a), built upon the European strategy for AI, it is underlined that in order to address the opportunities and challenges raised by AI systems and to achieve the objective of trustworthy, ethical and human-centric AI, it is necessary to rely on European values and to ensure 'that new technologies are at the service of all Europeans – improving their lives while respecting their rights' (COM, 2020a). In the same direction, a coordinated European approach on the human and ethical implications of AI, as well as a reflection on the better use of Big Data for innovation, was announced in her political guidelines by the Commission President Ursula von der Leyen (2019).

BIG DATA AND ITS IMPACT ON PRIVACY

Human Dignity at Risk Due to the 'Creep Factor' of Big Data

The use of Big Data, new surveillance tools and data gathering techniques represent a fundamental step for the European economy. Nevertheless, it also poses significant legal problems from a data protection perspective, despite the renewed legal framework (General Regulation on the Protection of Personal Data, GDPR). In the Big Data paradigm, traditional methods and notions of privacy protections might be inadequate in some instances (e.g. informed consent approaches), whilst the data are often used and re-used in ways that were inconceivable when the data were collected.

As acknowledged by the EDPS, the respect for human dignity is strictly interrelated with the respect for the right to privacy and the right to the protection of personal data. That human dignity is an inviolable right of human beings is recognised in the European Charter of Fundamental Rights. This essential right might be infringed by violations like objectification, which occurs when an individual is treated as an object serving someone else's purposes (European Data Protection Supervisor, Opinion 4/2015).

The impact of Big Data technologies on privacy (and thereby human dignity) ranges from group privacy and high-tech profiling, to data discrimination and automated decision making. It is even more significant if people disseminate personal data in the digital world at different levels of awareness throughout their main life phases. Here, people can often make themselves almost completely transparent for data miners who use freely accessible data from social networks and other data associated with an IP address for profiling purposes.

This 'creep factor' of Big Data, due to unethical and deliberate practices, bypasses the intent of privacy law. Such practices are allowed by advances in analysing and using Big Data for revealing previously private individual data (or statistically close proxies for it) and often have the final aim of targeting and profiling customers.

Another concern in relation to Big Data is the possibility of the re-identification of the data subject after the process of anonymisation. This might occur using technologies of de-anonymisation made available by the increased computational power of modern day personal computers, enabling a trace back to the original personal data. Indeed, traditional anonymisation techniques, making each data entry non-identifiable by removing (or substituting) uniquely identifiable information, has limits: despite the substitution of users' personal information in a dataset, de-anonymisation can be overcome in a relatively short period of time through simple links between such anonymous datasets, other datasets (e.g. web search history) and personal data. Re-identification of the data subject might also derive from the powerful insights produced when multiple and specific datasets from different sources are joined. This might allow interested parties to uniquely identify specific physical persons or small groups of persons, with varying degrees of certainty.

The re-identification of data poses serious privacy concerns: once anonymised (or pseudo-anonymised), data may be freely processed without any prior consent by the data subject, before the subject is then re-identified. The situation is exacerbated by the lack of adequate transparency regarding the use of Big Data: this affects the ability of a data subject to allow disclosure of his/her information and to control access to these data by third parties, also impacting civil rights.

It is advisable that organisations willing to use Big Data adopt transparent procedures and ensure that these procedures are easily accessible and knowable by the public. In this way, an ethical perspective would truly drive innovation and boundary setting, properly taking into account the individual's need for privacy and self-determination.

New Types of Stigmatisation and Manipulation of Civil Rights in the 'Group Privacy' Landscape

The right to privacy is undergoing an evolution. Originally arising as the right to be let alone and to exclude others from personal facts, over the years it has shifted to the right to being able to control personal data, and is now moving further in the direction of improved control. The current direction is towards the right to manage identity and the analytical profile created by third parties which select the relevant patterns to be considered in metadata. This third phase dwells not only on data that enable the identification of specific physical persons, but more on data suitable for finding out specific patterns of behaviour such as health data, shopping preferences, health status, sleep cycles, mobility patterns, online consumption, friendships, etc., of groups rather than of individuals. Despite the data being anonymous (in the sense of being de-individualised), groups are increasingly becoming more transparent: indeed, stripping data from all elements pertaining to any sort of group belongingness would result in stripping the collection itself from its content and therefore its usefulness.

This information gathered from Big Data can be used in a targeted way to encourage people to behave or consume in a certain way. Targeted marketing is an example, but other initiatives (for instance, in the political landscape), based on the ability of Big Data to discover hidden correlations and on the inferred preferences and conditions of a specific group, could be adopted to encourage or discourage a certain behaviour, with incentives whose purposes are less transparent (including not only market intelligence, but other forms of manipulations in several sectors – such as in voting behaviour).

New types of stigmatisation might also arise, for instance, in relation to the commercial choices and other personal information of groups. Forms of discrimination are likely, especially when the groups get smaller (identified by geographical, age, sex, etc. settings). In this sense, Big Data techniques might eclipse longstanding civil rights protections.

What increases ethics concern is the related collection and aggregation of mass Big Data, and the resulting structured information and quantitative analysis for this purpose that are not subject to the application of current data protection regulations. Therefore, innovative ways of re-thinking citizens' protection are needed, capable of offering adequate and full protection.

The 'Sharing the Wealth' Model and the 'Personal Data Store' Approach for Balancing Big Data Exploitation and Data Protection

As pointed out by the EU Agency for Network and Information Security (ENISA), it is necessary to overcome the conceptual conflict between privacy and Big Data and between privacy and innovation. The need is to shift '… the discussion from "big data versus privacy" to "big data with privacy"', and to recognise the privacy and data protection principles as 'an essential value of big data, not only for the benefit of the individuals, but also for the very prosperity of big data analytics' (ENISA, 2015, p. 5). There is no dichotomy between ethics and innovation if feasible balancing solutions are figured out and implemented. The respect for citizens' privacy and dignity and the exploitation of Big Data's potential can fruitfully coexist and prosper together, balancing the fundamental human values (privacy, confidentiality, transparency, identity, free choice and others) with the compelling uses of Big Data for economic gains. This is aligned with EDPS's recent opinion (European Data Protection Supervisor, Opinion 3/2020 on the European strategy for data) underlining that data strategy's objectives could encompass 'to prove the viability and sustainability of an alternative data economy model – open, fair and democratic' where, in contrast with the current predominant business model,

> characterised by unprecedented concentration of data in a handful of powerful players, as well as pervasive tracking, the European data space should serve as an example of transparency, effective accountability and proper balance between the interests of the individual data subjects and the shared interest of the society as a whole.

The key question is how to ensure this coexistence and the underlying balance is achieved. The answer is not simple and relies on multiple dimensions. From a technological perspective, Privacy by Design and Privacy Enhancing Technologies (PETs) come into play.[4]

As stated by the EU Regulation 2016/679, the data protection principles should be taken into consideration at a very early stage, as well as privacy measures and PETs should be identified in conjunction with the determination of the means for processing and deployed at the time of the processing itself. ENISA proposed an array of privacy by design strategies, ranging from data minimisation and separate processing of personal data, to hiding personal data and their interrelation, opting for the highest level of aggregation. The PETs to implement these strategies are already applied in the Big Data industry: they rely on anonymisation, encryption, transparency and access, security and accountability control, consent ownership and control mechanisms. Even so, an adequate investment in this sector is required, as confirmed by the small number of patents for PETs compared to those granted for data analytics technologies. Efforts need to be directed towards strengthening data subject control thereby bringing transparency and trust in the online environment. In fact, trust has emerged as a complex topic within the contemporary Big Data landscape. At the same time, it has become a key factor for economic development and for the adoption of new services, such as public e-government services, as well as for users' acceptance to provide personal data. In some instances, such as in the medical field, the choice not to provide a full disclosure of the requested information might impact the individual's wellbeing or health (besides indirectly hindering progress in research), given that these are personal data and the trust relationship with the data collector (e.g. the staff of a hospital) is functional to the individual's wellbeing and/health.

The 'sharing the wealth' strategy proposed by Tene and Polonetsky (2013) for addressing Big Data challenges is based on the idea of providing individuals access to their data in a usable format and, above all, allowing them to take advantage of solutions capable of analysing their own data and drawing useful conclusions from it. The underlying vision is to share the wealth individuals' data helps to create with individuals themselves, letting them make use of and benefit from their own personal data. This approach is also aligned with the vision of the Big Data Value Association (BDVA Position Paper, 2019), which outlines opportunities of data economy arising over the next decade for the industry (business), the private users (citizens as customers), the research and academic community (science) and local, national and European government and public bodies (government).

Other authors (Rubinstein, 2013) underline the potentialities of a new business model based on the personal data store or personal data space (PDS). Such a business model shifts data acquisition and control to a user-centric paradigm, based on better control of data and joint benefits from its use. This solution (and the necessary implementing technology), if developed, might enable users' empowerment and full control over their personal data. In fact, it would permit users to gather, store, update, correct, analyse and/or share personal data, as well as having the ability to grant and withdraw consent to third parties for access to data. In this way, it would also work towards more accountable companies, where the commitment in personal data protection might become an economic asset for digital players.

PDS are also aligned with the importance of data portability, strongly advocated by the EDPS in view of guaranteeing people the right to access, control and correct their personal data, whilst enhancing their awareness. Data portability also nurtures the suggested approach of allowing people to share the benefits of data and can foster the development of a more competitive market environment, where the data protection policy is transformed into a *strategical economic asset*, thus triggering a virtuous circle. Companies would be encouraged to invest to find and implement the best ways to guarantee the privacy of their customers: indeed, data portability allows customers to switch providers more easily, also by taking into account the provider more committed to respecting personal data and to investing in privacy-friendly technical measures and internal procedures.

The 'sharing the wealth' paradigm and the potentialities of a new ethically driven business model relying on personal data are at the basis of the European Project DataVaults – 'Persistent Personal DataVaults Empowering a Secure and Privacy Preserving Data Storage, Analysis, Sharing and Monetisation Platform' (Grant Agreement no. 871755), funded under the H2020 Programme.[5] This project, currently under development, is aimed at setting, sustaining and mobilising an ever-growing ecosystem for personal data and insights sharing, capable of enhancing the collaboration between stakeholders (data owners and data seekers). Its value-driven tools and methods for addressing concerns about privacy, data protection, security and Intellectual Property Rights (IPR) ownership will enable the ethically sound sharing both of personal data and proprietary/commercial/industrial data, following strict and fair mechanism for defining how to generate, capture, release and cash out value for the benefit of all the stakeholders involved, as well as securing value flow based on smart contract, moving towards a win–win data sharing ecosystem.

The European Privacy Association even proposes to see data protection for digital companies not as mere legal compliance obligations, but as part of a broader corporate social responsibility) and socially responsible investments in the Big Data industry. It is recommended to valorise them as assets within renewed business models, able to help companies responsibly achieve their economic targets.

From a wider perspective, as also underlined by BDVA (2020) in particular in relation to the Smart Manufacturing environment, the soft law in the form of codes of conduct could bring a set of advantages at ecosystem level in each domain. In fact, such sources are expected to offer guidance and to address in meaningful, flexible and practical ways the immediate issues and ethical challenges of Big Data and AI innovations in each sector, going beyond current gaps in the legal system: they can operate as a rulebook, providing more granular ethical guidance as regards problems and concerns, resulting in an increase of confidence and legal certainty of individuals which also encompass trust building and consolidation.

In parallel, this calls for promoting the acquisition of skills on privacy as a value and right, on ethical issues of behaviour profiling, ownership of personal contents, virtual identity-related risks and digital reputation control, as well as on other topics related to Big Data advancements. On this purpose, Bachelor's and Master's degree programmes in Data Science, Informatics, Computer Science, Artificial Intelligence and related subjects could be adequately integrated in order to cover these themes. In this way, human resources in Big Data businesses could include ad hoc professional figures.

At the same time, in order to promote the commitment of the business world, it is advisable that the efforts of those companies which invest in ethical relationships with customers are recognised by governments and properly communicated by the companies themselves to their customer base. The certification approach should also be explored, as inspired by the Ethics Certification Program for Autonomous and Intelligent Systems launched by the Institute of Electrical and Electronics Engineers for AIS[6] products, services and systems.

This would let them further benefit in terms of improved reputation and let them increase the trust of customers towards their products and services. At the same time, information on business ethics violations occurring through the improper use of Big Data analytics should be transparent and not kept opaque to consumers.

A Critical Perspective on the 'Notice and Consent' Model and on the Role of Transparency in the Evolving World of Big Data Analytics

Emerging commentators argue that the data protection principles, as embodied in national and EU law, are no longer adequate to deal with the Big Data world: in particular, they criticise the role of transparency in the evolving world of Big Data analytics, assuming that it no longer makes sense considering the complex and opaque nature of algorithms. They also debate the actual suitability of the so-called 'notice and consent' model, on the grounds of consumers' lack of time, willingness or ability to read long privacy notices.

Others prefer to emphasise accountability, as opposed to transparency for answering Big Data ethics challenges, being focussed on mechanisms more aligned with the nature of Big Data (such as assessing the technical design of algorithms and auditability). GDPR itself highlights, besides the role of transparency, the growing importance of accountability.

Instead of denying the role of transparency in the Big Data context, others suggest that it is not possible to offer a wholesale replacement for transparency and propose a more 'layered' approach to it (for instance, as regards privacy notices to individuals and also the information detail), in conjunction with a greater level of detail and access being given to auditors and accredited certification bodies.

On the contrary, transparency itself might be considered as a requirement needed for accountability and seems unavoidable in the context of respect for human dignity. Traditional notice and consent models might be rather insufficient and obsolete in view of the effective exercise of control and in order to avoid a situation where individuals feel powerless in relation to their data. Nevertheless, to overcome this weakness, an alternative, more challenging path is to make consent more granular and capable of covering all the different processing (and related) purposes and the re-use of personal data. This effort should be combined with increased citizens' awareness and a higher participation level, as well as with effective solutions to guarantee the so-called right to be forgotten.

In the same user-centric approach, based on control and joint benefits and promoted by EC and European-wide initiatives,[7] a number of views foster new approaches premised on consumer empowerment in the data-driven business world. These approaches strongly aligned with the transparency and accountability requirements, ask for proper internal policies and control systems, focussed on pragmatic, smart and dynamic solutions and able to prevent the risk of companies becoming stuck in bureaucracy.

DISCRIMINATION, SOCIAL COOLING, BIG DATA DIVIDE AND SOCIAL SORTING

A possible side effect of datafication is the potential risk of discrimination of data mining technologies in several aspects of daily life, such as employment and credit scoring (Favaretto, De Clercq, & Elger, 2019). It ranges from discriminatory practices based on profiling and related privacy concerns (e.g. racial profiling enabled by Big Data platforms in subtle ways by targeting characteristics like home address and misleading vulnerable less-educated groups with scams of harmful offers),[8] to the impact of Big Data in the context of the daily operation of organisations and public administrations (e.g. within human resources offices). In the latter context, crucial decisions, like those about employment, might rely on the use of Big Data practices which might bring the risk of unfair treatment through discrimination based on gender, race, disability, national origin, sexual orientation and so on.

Social Cooling as a Side Effect of Big Data

We live in a society where everything can be given a score and critical life changing opportunities are increasingly determined by such scoring systems, often obtained through secret predictive algorithms applied to data to determine which individuals or which social group has value. It is therefore essential to consider human values as oversight in the design and implementation of these systems and, at the same time, to guarantee that the policies and practices using data and scoring machines to make decisions are realised in a legal and ethical manner (including avoiding automated decision-making practices not compliant with regulatory boundaries set forth by art. 22 GDPR). Fair and accurate scoring systems have to be ensured, whilst also avoiding the risk that data might be biased to arbitrarily assign individuals to a stigmatising group. Such an assignment might potentially allow that decisions relevant for them are not fair and, in the end, might negatively affect their concrete opportunities.

Any Big Data system has to ensure that, if existing, automated decision making, especially in areas such as employment, health care, education and financial lending, operates fairly for all communities, and safeguards the interests of those who are disadvantaged. The use of Big Data, in other words, should not result in infringements of the fundamental rights of individuals, neither in differential treatment or indirect discrimination against groups of people, for instance, as regards the fairness and equality of opportunities for access to services.

As indicated by the European Parliament, all measures possible need to be taken to minimise algorithmic discrimination and bias and to develop a common ethical framework for the transparent processing of personal data and automated decision making. This common framework should guide data usage and the ongoing enforcement of EU law. From this perspective, it is necessary that the use of algorithms to provide services – useful for identifying patterns in data – rely on a comprehensive understanding of the context in which they are expected to function and are capable of picking up what matters. It is also essential to establish oversight activities and human intervention in automated systems as well, besides considering that Big Data needs to be coupled with room for politics and with mechanisms to hold power to account. In this way, unintended negative societal consequences of possible errors introduced by algorithms, especially in terms of the risk of systematic discrimination across society in the provision of services, might be prevented or at least minimised.

This will also limit the widening of one of the chilling effects of Big Data related to discrimination, the so-called social cooling. Social cooling could limit people's desire to take risks or exercise free speech, which, over the long term, could 'cool down' society.[9] The term describes the long-term negative side effects in terms, for instance, of self-censorship, risk-aversion and exercise of free speech, of living in a reputation economy where Big Data practices that lack an ethical dimension are increasingly apparent and intrusive.

Social cooling is due to people's emerging perception that their data, including the data reflecting their weaknesses, is turned into thousands of different scores and that their resulting 'digital reputation' could limit their opportunities. As a consequence, they feel pressure to conform to a bureaucratic average, start to apply self-censorship and tend to change their behaviour to achieve better scores. This might result, especially if public awareness remains very low, in increased social rigidity, limiting people's ability and willingness to protest injustice and, in the end, in a subtle form of socio-political control. The related societal question is whether this trend will have an impact on the human ability to evolve as a society, where minority views are still able to flourish.

The social cooling effect emphasises another dimension of a mature and nuanced perception of data and privacy: its ability to protect the right to be imperfect, in other words the right to be human.

Big Data Divide

The expression Big Data Divide has a two-fold meaning. First, it refers to the difficulty in accessing services delivered through the use of the Internet and other new technologies and to the complexity in understanding how these technologies and related services work. This kind of digital divide might have consequences, for instance, with regard to online job hunting: senior citizens, who are unfamiliar with this new way of job hunting, can be harmed in terms of lost job opportunities. The same may happen with regard to other tools such as online dating services for finding a new partner or for social interactions. The consequences might be frustration and social withdrawal. Similarly, inclusion concerns are related to the possible definition of new policies based on a data-driven approach (e.g. data collected via sensors, social media, etc.); there is the concrete possibility that some individuals or portions of a society might not be considered. The risk is that the new policy will only take into account the needs of people having access to the given technological means. Secondly, the notion of a 'Big Data divide' refers to the asymmetric relationship between those 'who collect, store, and mine large quantities of data, and those whom data collection targets' (Andrejevic, 2014). The Big Data divide is perceived as potentially able to exacerbate power imbalances in the digital era and increase the individual's sense of powerlessness in relation to emerging forms of data collection and data mining.

Furthermore, it has been argued that Big Data and data mining emphasise correlation and prediction and call to mind the emergent Big Data-driven forms of social sorting (and related risk of discrimination). This remark refers to the ability – enabled by Big Data and data mining – of discerning unexpected, unanticipated correlations and of generating patterns of actionable information. Such ability provides powerful insights for decision making and prediction purposes, unavailable to those without access to such data, processing power and findings: those with access are advantageously positioned compared to those without it.

Predictive analytics for data-driven decision making and social sorting can also lead to 'predictive policing' (Meijer & Wessels, 2019), where extra surveillance is set for certain individuals, groups or streets if it is more likely that a crime can be committed. Though systematic empirical research, capable of generating an evidence base on the benefits and drawbacks of this practice, seems to be still missing, the predictive policy encompasses a political challenge: if it is difficult to ignore these kinds of findings and doing nothing to prevent the occurrence of the crime, at the same time the risk of stigmatisation of such individuals or groups has to be tackled. A balance could be sought considering, for instance, the intervention threshold and correlating the type of intervention with the likelihood of crime anticipated by the algorithms, being careful to exclude incidental co-occurrences.

Big Data from the Public Sector Perspective

Big Data for Public Use

Another area to investigate is how Big Data might be used for public good and with public support.

Both in the 'European Strategy for Data' (COM, 2020b) and in the recent Proposal for a Regulation on European Data Governance ('Data Governance Act') which is the first of a set of measures announced in the strategy, data altruism is facilitated, meaning 'data voluntarily made available by individuals or companies for the common good' (COM, 2020c). The increasing amounts of data in society might change the type of evidence that is available for policy makers and, at the same time, policy makers can linger over computer models and predictive analytics as a basis for their decisions. The chance to draw meaningful insights (relevant for policy elaboration purposes) from data would require a comprehensive data infrastructure, where data sources are well organised and can be accessed by authorised people for the appropriate use. The discussion mainly explores the opportunities in local services in view of accompanying local decisions by evidence for securing investment from central budget holders. The surveys ranged from identifying what approaches work better for the public at a lower cost to efficaciously demonstrate and show where resources are lacking and investment needed. However, the possible use of data analysis in many local authorities is being confronted by more traditional approaches, as well as with civil servants' diffidence in exploiting the potentialities of cutting-edge technologies. Thereby an organisational and cultural change needs to be supported, through awareness campaigns and other initiatives.

An interesting example of how Big Data can be exploited for the common good and public interest in conjunction with private business' priorities is the solution developed in the project AEGIS – 'Advanced Big Data Value Chain for Public Safety and Personal Security' (Grant Agreement no. 732189), funded by the European Commission in the H2020 Programme. The project brought

> together the data, the network and the technologies to create a curated, semantically enhanced, interlinked and multilingual repository for public and personal safety-related Big Data. It delivers a data-driven innovation that expands over multiple business sectors and takes into consideration structured, unstructured and multilingual datasets, rejuvenates existing models and facilitates organisations in the Public Safety and Personal Security linked sectors to provide better & personalised services to their users.[10]

The services enabled by this technology aim to generate value from Big Data and renovate the Public Safety and Personal Security sector, positively influencing the welfare and protection of the general public. Project achievements aim to have positive impacts in terms of economic growth and enhanced public security, as well as for individuals, by improving safety and wellbeing through prevention and protection from dangers affecting safety (such as accidents or disasters).

Dataveillance, Big Data Governance and Legislation

Big Data poses multiple strategic challenges for governance and legislation, with the final aim of minimising harm and maximising benefit from the use of data. Such challenges require consideration of risks and risk management.

The first issue is related to the practice of the so-called ‘dataveillance’, where the use of data improves surveillance and security. It refers to the continuous monitoring and collecting of users’ online data (data resulting from email, credit card transactions, GPS coordinates, social networks, etc.), including communication and other actions across various platforms and digital media, as well as metadata. This kind of surveillance is partially unknown and happens discreetly. Dataveillance can be individual dataveillance (concerning the individual’s personal data), mass dataveillance (concerning data on groups of people) and facilitative mechanisms (without either considering the individual as part of a group, or targeting any specific group).

In the public perception, the idea that one’s position and activity might be in some way tracked at most times has become an ordinary fact of life, in conjunction with an increased perception of safety: almost everyone is aware of the ubiquitous use of CCTV[11] circuits, the GPS[12] positioning capabilities inside mobile devices, the use of credit cards and ATM[13] cards and other forms of tracking. On the contrary, this active surveillance might also have an impact on citizens’ liberties and might be used by governments (and businesses too) for unethical purposes.

Ethical concerns revolve around individual rights and liberties, as well as on the ‘data trust deficit’, whereby citizens have lower levels of trust in institutions to use their data appropriately.

Other important tools for accountability to the public should be implemented, in order to avoid the public perception that there are no mechanisms for accountability outside of public outcry. This implies tackling the challenge for Big Data governance. For instance, it would be useful if there were a formulation and upholding of an authoritative ethical framework at the national or international level, drawing upon a wide range of knowledge, skills and interests across the public, private and academic sectors, and confirmed by a wide public consultation.

Alongside this ethical framework an update of the current legislative system would be opportune for minimising harm and maximising benefit from the use of data: in fact, the regulation is developing at a much slower pace than the Big Data technology and its applications. This results in the business community’s responsibility to decide how to bridle the insights offered by data from the multiple data sources and devices, according to their respective core ethical values.

DATA OWNERSHIP

Another dimension of the debate on Big Data also revolves around data ownership, which might be considered as a sort of IPR issue separate from technology IPR.

The latter refers to the procedures and technologies used to acquire, process, curate, analyse and use the data. Big Data technology IPRs are mostly covered by the general considerations applicable for software and hardware IPRs and the related business processes, though considered in the Big Data domain. In this view, special IPR approaches are not needed, being covered by existing models and approaches existing for the assertion, assignment and enforcement of copyright, design rights, trademarks and patents for IT technology in general.

On the contrary, data ownership refers to the IP related to the substantive data itself, including both raw data and derived data. The main IP rights in relation to data are database rights, copyright and confidentiality: due to the fact that database rights and copyright protect expression and form rather than the substance of information, the best form of IP protection for data is often considered the one offered by the provisions safeguarding the confidentiality of information, being capable of protecting the substance of data that is not generally publicly known.

IP challenges in the Big Data domain are different from existing approaches and need special care, especially as regards protection, security and liability, besides data ownership. At the same time, addressing the challenges raised by IP issues is essential, considering the expected high incomes due to increased Big Data innovation and technology diffusion.

Data ownership and the rights to use data might be covered by copyright and related contracts which are valid when collecting the data, often including also confidentiality clauses. In case of further processing of big datasets, it has to be explored when and how this creates new ownership: in fact, the acquisition of data, its curation and combination with other datasets, as well as possible analysis of them and resulting insights, creates new rights to the resulting data, which need be asserted and enforced.

Regardless of the considerations stemming from the regulatory perspective, notably Directive 96/9/EC on the legal protection of databases, the main ethical dilemma concerns how to consider user's data. In other words, the question is to whom these data belong: still to the user, or to the company that conducted the analyses, or the company that gathered the original data?[14]

All these issues should not only be specifically addressed by national and European legislation on IPR in relation to data, which is of uncertain scope at the moment, but also investigated by the data ethics debate: best practices for collection, recommendations and guidelines would be very useful. Currently, a key role for addressing this issues is played by contract provisions.

In view of ensuring the fair attribution of value represented in data creation, but, at the same time, considering the multiple, competing interests at stake in B2B[15] data sharing, balancing operations should be conducted between the data producers' interest to remain in control of their data and to retain their rights as the original owners, the public interest in avoiding data monopolies (due to the fact that data still fuel innovation, creativity and research) and data subjects' interest in their personal information collected by a company.

Regarding the first of these interests and the related ownership claims, the legal framework is still uncertain and fragmented. The situation is further complicated by the difficulty of applying legal categories: the data are an intangible good difficult to define and the same legal concept of data ownership is not clearly defined. Many questions arise, such as: does existing EU law provide sufficient protection for data? If not, what more is needed? Are data capable of ownership (sui generis right or copyright law)? Is there a legal basis for claims of ownership of data? Is there the need of enactment of exclusive rights in data? Or is it better to explore alternatives?

Regarding alternatives, an interesting option is to provide the factual exclusivity of data through flexible and pragmatic solutions able to provide certainty and predictability, by combining agile contracting with enabling technological tools. As for the contractual layer of this solution, it consists of ad hoc and on-the-fly B2B data exchange contracts, provided under the well-defined data sovereignty principle to safeguard data producers' control over data generated. For this purpose, access and usage policies or protocols need to be implemented. At the same time, it is necessary to establish a trade-off with other interests, like individual 'interest' over personal data, in this case. On the contrary, the technological layer provides enabling technologies to implement and enforce the terms and conditions set forth by the data sharing agreements. Technologies to be explored include, for instance, sticky policies, Blockchain, Distributed Ledger Technologies and smart contract, Digital Rights Management technologies and APIs.[16]

This kind of solution is well-developed by the International Data Space Association (IDSA),[17] consisting of more than one hundred companies and institutions from various industries and of different sizes from 20 countries collaborating to design and develop a trustworthy architecture for the data economy. Its vision and reference architecture rotate around the concept of 'data sovereignty', defined as 'a natural person's or corporate entity's capability of being entirely self-determined with regard to its data' (IDSA, 2019). Data sovereignty, which is materialised in 'terms and conditions' (such as time to live, forwarding rights, pricing information, etc.) linked to data before it is exchanged and shared. Such terms and conditions are supported and enforced through the technical infrastructure, including tools for the secure and trusted authorisation, authentication and data exchange (such as blockchain, smart contracts, identity management, point-to-point encryption, etc.) to be customised to the needs of individual participants.

In line with the joint benefit approach and with the related user-centric business model based on PDS, a similar path could be further extended also for strengthening the contract provisions underpinning high-value personal data ecosystems leaving the process under the individuals' control, like in the DataVaults Project. This is also the goal of the new Smart Cities Marketplace Initiative within the Citizen Focus Action Cluster: 'Citizen Control of Personal Data',[18] launched on 27 January 2021. Its intention is

> to contribute to speeding up the adoption, at scale, of common open urban data platforms, and ensure that 300 million European citizens are served by cities with competent urban data platforms, by 2025. The potential for citizen's personal data to contribute to data ecosystems will be significantly enhanced by introducing secure, ethical and legal access to this highly coveted and valuable personal data, incorporating citizen-generated data as 'city data'.

Novel contract rights, including IPR provisions, might be further spread in the data-driven economy, in view of confirming users' control over their data, as well as their empowerment, thereby contributing to going beyond possible existing differences between national laws and gaps in the European legislation.

Nevertheless, as in the past, when the IPR development has followed the commercialising of innovation, the growth of the Big Data market is likely to generate also the further renewal of the IPRs' regulatory framework underpinning it and to pave the way to set a coherent system at European level.

CONCLUSIONS

The rise of Big Data and the underlying ability to capture and analyse datasets from highly diversified contexts and generate novel, unanticipated knowledge, as well as AI developments relying on data, are capable of producing economic growth and bringing relevant benefits, both at the social and the individual level. This rapidly sprawling phenomenon is expected to have significant influence on governance, policing, economics, security, science, education, health care and much more.

The collection of Big Data and inferences based on them are sources enabling both economic growth and generation of value, with the potential to bring further improvement to everyday life in the near future. The examples span from road safety, to health services, agriculture, retail, education and climate change mitigation. Possible improvements rely on the direct use and collection of Big Data or on inferences or 'nowcasting' based on them: new knowledge and insights are generated, as well as real-time reports and analyses with alerting purposes can be produced.

At the same time, Big Data practices and techniques put at stake several ethical, social and policy challenges, threats and potential hurdles. They are often interrelated and range from concerns related to data ownership to the 'datafication' of society, to privacy dilemmas and the potential trade-off between privacy and data analytics progress, social cooling, dataveillance, discriminatory practices and the emerging Big Data divide. Such challenges, threats and potential hurdles also include, for instance, the data-driven business ethics violations, the 'data trust deficit', the concerns due to the use of Big Data in the public sector and the desirable role of the government towards the fair policy development and the provision of enhanced public services.

These and similar items need greater ethics engagement and reflection, in the framework of an interdependent ecosystem, composed of different and complementary competences (primarily legislators, data-driven businesses, IT developers and data scientists, civil society organisations and academia) in order to come up with a Big Data market fully respectful of human dignity and citizens' rights and susceptible of further development in an ethically acceptable way.

The fruitful development of this ecosystem might also require the adjustment of familiar conceptual models and archetypes of research ethics, to better align them with the epistemic conditions of Big Data and the data analytics work. The envisioned alignment should reflect also on the shift towards algorithmic knowledge production to identify and address eventual mismatches between the Big

Data research and the extant research ethics regimes. In parallel, inquiry should be moved away from considering only traditional categories of harm (e.g. physical pain and psychological distress) to cover other types and forms (e.g. effects of the perennial surveillance on human behaviour and dignity and group discrimination). Likewise, the concept of the human subject and related foundational assumptions should be revisited to include not only individuals, but also distributed groupings or classifications.

The need to productively re-think some concepts of research ethics and regulations, due to the development of large-scale data analytics, represents an opportunity to reaffirm basic principles and values of human dignity, respect, transparency, accountability and justice. The final aim is to contribute to shaping the future trajectory of the Big Data revolution, with its interplay with AI breakthroughs, in a way that is truly responsive to foundational ethical principles.

NOTES

1. COM (2020b). This communication is part of a wider package of strategic documents, including the COM (2020a), the Communication on Shaping Europe's digital future.
2. The volume of data produced is growing quickly, from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025 in the world (IDC, 2018).
3. European Data Protection Supervisor, Opinion 3/2020 on the European strategy for data. In the same document, the EDPS applauds the EC's commitment to safeguard that European fundamental rights and values, underpinning all aspects of the data strategy and its implementation.
4. A high-value description and classification of the PETs and their role was provided by the e-SIDES project (https://e-sides.eu/e-sides-project) deliverables. In the e-SIDES Deliverable D3.2 and in the related White Paper, the overview of existing PETs is accompanied by an assessment methodology of them for facing legal and ethical implications based on interviews and desk-research: it provides, on the one hand, the technology-specific assessment of selected classes of PETs, and, on the other hand, a more general assessment of such technologies.
5. In particular within the call H2020-ICT-2019-2, topic ICT-13-2018-2019 'Supporting the emergence of data markets and the data economy'. Further information on DataVaults can be retrieved at the following link: https://www.datavaults.eu/.
6. Autonomous and Intelligent Systems.
7. See, for instance, EC's COM (2019) and EFFRA (2013, 2020).
8. An interesting reading on the risk of racial profiling which might be generated by new technological tools and methods, such as Big Data, automated decision making and AI is the 'General recommendation No. 36 (2020) on preventing and combating racial profiling by law enforcement officials' released by the United Nations' Committee on the Elimination of Racial Discrimination (2020) on 17 December.
9. https://www.socialcooling.com/
10. https://cordis.europa.eu/project/rcn/206179_it.html
11. Closed-Circuit Television.
12. Global Positioning System.
13. Automated Teller Machine.
14. An interesting reading on this topic is AA.VV (2016).
15. Business to Business
16. Application Programming Interfaces.
17. https://www.internationaldataspaces.org/
18. https://smart-cities-marketplace.ec.europa.eu/news/new-initiative-citizen-control-personal-data-within-citizen-focus-action-cluster. This initiative is committed to seek to

remove existing constrains and helping to create the conditions and relationships whereby 'the citizen will be willing to share personal data with a city and with other actors in the data economy. The ambition behind this new initiative is to give the smart cities movement a boost by providing cities with access to a rich personal data pool. This pool of data, in turn, would stimulate further activity within the data economy, accelerate the take-up of urban data platforms and contribute to the improvement of mobility, health, energy efficiency and better governance among other'.

REFERENCES

AA.VV. (2016). Data ownership and access to data. Position statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the current European debate.

Andrejevic, M. (2014). The Big Data divide. *International Journal of Communication*, *8*, 1673–1689.

Bailly, M., & Manyika, J. (2013). *Is manufacturing 'cool' again*?. McKinsey Global Institute. Retrieved from https://www.brookings.edu/opinions/is-manufacturing-cool-again/

BDVA. (2020). Big Data challenges in smart manufacturing industry. A White paper on digital Europe Big Data challenges for smart manufacturing industry. Retrieved from https://bdva.eu/sites/default/files/BDVA_SMI_Discussion_Paper_Web_Version.pdf. Accessed on July 26, 2021.

BDVA Position Paper. (2019, April) Towards a European data sharing space – Enabling data exchange and unlocking AI potential. Retrieved from https://www.bdva.eu/node/1277. Accessed on July 26, 2021.

COM. (2019). 168 final "Building trust in human-centric artificial intelligence". Retrieved from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52019DC0168. Accessed on July 26, 2021.

COM. (2020a). 65 final "White paper on artificial intelligence – A European approach to excellence and trust". Retrieved from https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Accessed on July 26, 2021.

COM. (2020b). 66 final "A European strategy for data". Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066. Accessed on July 26, 2021.

COM. (2020c). 767 final "Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act)". Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767. Accessed on July 26, 2021.

DataVaults Project. Retrieved from https://www.datavaults.eu/

EFFRA. (2013). Factories of the future multi-annual roadmap for the contractual PPP under Horizon 2020. Retrieved from https://www.effra.eu/sites/default/files/190312_effra_roadmapmanufacturingppp_eversion.pdf

EFFRA. (2020). Vision for a manufacturing partnership in Horizon Europe 2021–2027.

ENISA. (2015). Privacy by design in Big Data. An overview of privacy enhancing technologies in the era of Big Data analytics. Retrieved from www.enisa.europa.eu. Accessed on July 26, 2021.

e-SIDES Deliverable D3.2 and White Paper. (2018). How effective are privacy-enhancing technologies in addressing ethical and societal issues?. Retrieved from https://e-sides.eu/resources. Accessed on July 26, 2021.

e-SIDES Project. Retrieved from https://e-sides.eu/e-sides-project

European Data Protection Supervisor, Opinion 3/2020 on the European strategy for data. Retrieved from https://edps.europa.eu/sites/default/files/publication/20-06-16_opinion_data_strategy_en.pdf

European Data Protection Supervisor, Opinion 4/2015. Towards a new digital ethics. Data, dignity and technology, 11 September 2015. Retrieved from https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf

Executive Office of the President. (2014). Big Data: Seizing opportunities, preserving values. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. Accessed July 26, 2021.

Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: Perils, promises and solutions. A systematic review. *Journal of Big Data*, *6*(1), 12.

IDSA. (2019). Reference architecture model, version 3.0. Retrieved from https://internationaldataspaces.org/use/reference-architecture/ Accessed on July 26, 2021.

Kagermann, H., & Wahlster, W. (2013). *Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative Industrie 4.0*. Final report of the Industrie 4.0 Working Group, AcatechVNational Academy of Science and Engineering, Germany. Retrieved from https://www.academia.edu/36867338/Securing_the_future_of_German_manufacturing_industry_Recommendations_for_implementing_the_strategic_initiative_INDUSTRIE_4_0_Final_report_of_the_Industrie_4_0_Working_Group. Accessed on July 26, 2021.

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, *42*(12), 1031–1039.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, *3*(1), 1–14.

Reichert, P. (2014). Comarch EDI platform case study: The advanced electronic data interchange hub as a supply-chain performance booster. In P. Golinska (Ed.), *Logistics operations, supply chain management and sustainability* (pp.143–155). Cham: Springer.

Reinsel, R., Gantz, J., Rydning, J. (2018). *Data Age 2025. The digitization of the World. From Edge to Core. An IDC White Paper*. Retrieved at https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Rubinstein, I. S. (2013). Big Data: The end of privacy or a new beginning?. *International Data Privacy Law*, *3*(2), 74–87.

Smart Cities Marketplace Initiative within the Citizen Focus Action Cluster. (2021). Citizen control of personal data. Retrieved from https://smart-cities-marketplace.ec.europa.eu/news/new-initiative-citizen-control-personal-data-within-citizen-focus-action-cluster

Tene, O., & Polonetsky, J. (2013). Big Data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5), 237–273.

United Nations' Committee on the Elimination of Racial Discrimination. (2020). General recommendation No. 36 on preventing and combating racial profiling by law enforcement officials. Retrieved from https://digitallibrary.un.org/record/3897913. Accessed on July 26, 2021.

van Brakel, R. (2016). Pre-emptive Big Data surveillance and its (dis)empowering consequences: The case of predictive policing. In B. van der Sloot, D. Broeders, E. Schrijvers (Eds.), *Exploring the Boundaries of Big Data* (pp.117-141). Amsterdam: Amsterdam University

von der Leyen, U. (2019). A union that strives for more. My agenda for Europe. Retrieved from https://ec.europa.eu/info/sites/info/files/political-guidelines-next-commission_en_0.pdf. Accessed on July 26, 2021.