

D2.1 Refined Research Framework



Ethical and Societal Implications of Data Sciences

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731873

This document does reflect the authors view only.

The European Commission is not responsible for any use that may be made of the information this document contains.

Copyright belongs to the authors of this document.

Use of any materials from this document should be referenced and is at the user's own risk.

1. Executive Summary

Data-driven innovation is deeply transforming society and the economy. Although there are potentially enormous economic and social benefits, this innovation also brings new challenges related to, for instance, individual and collective privacy, security democracy and participation. The main objective of the Coordination and Support Action e-SIDES is to complement the research on privacy-preserving big data technologies, by analysing, mapping and clearly identifying the main societal and ethical challenges emerging from the adoption of big data technologies, conforming to the principles of responsible research and innovation; setting up and organizing a sustainable dialogue between industry, research and social actors, as well as networking with the main Research and Innovation Actions and Large Scale Pilots and other framework programme projects interested in these issues.

The present deliverable serves as an initial research framework, laying foundations for the further activities within the project. The idea behind such framework is to obtain the common ground regarding the key concepts to be applied, the scope of the project and its main goals.

With this idea in mind, the present document is divided into sections corresponding to different goals pursued in the project and its focus areas. First, the general part describes the scope and objectives of the entire e-SIDES project and explains the key terminology to be applied throughout the project in order, on the one hand, to delineate the relevant subject matter of the undertaken research and, on the other, to ensure the coherence between different sections of the project. Second, the document explains the methodology and strategy for building the community of stakeholders who, through an ongoing dialog, will play a key role in informing the research activities undertaken within the project, enabling the 'big data on the ground' vision of this phenomenon. Third, the deliverable briefly explains the scope and focus of each substantive segment of the research agenda: the ethical, legal, societal and economic aspects of big data technologies, including the related initiatives ongoing in these fields in the EU and beyond.

Contents

1. Executive Summary.....	3
2. Introduction	5
2.1. The e-SIDES project: key questions.....	5
2.2. Structure of the deliverable and relationship to other deliverables	5
3. Conceptual framework and definitions	6
3.1. Scope.....	6
3.2. Goals	8
3.3. Terminology	9
4. Community building, stakeholders engagement.....	12
4.1. Stakeholder identification.....	13
4.2. Awareness raising	13
4.2.1. Community building and engagement.....	14
5. The ethical perspective	16
5.1. Introduction	16
5.2. Focus and related initiatives	17
5.3. Terminology	18
6. The legal perspective	20
6.1. Introduction	20
6.2. Focus and related initiatives	21
6.3. Terminology	23
7. The societal perspective	26
7.1. Introduction	26
7.2. Focus and related initiatives	26
7.3. Terminology	29
8. The economic perspective	32
8.1. Introduction	32
8.2. Focus and related initiatives	32
8.3. Terminology	33
9. Conclusions	35
10. Glossary.....	35
11. References	39

2. Introduction

2.1. The e-SIDES project: key questions

Data-driven innovation is deeply transforming society and the economy. Although there are potentially enormous economic and social benefits this innovation also brings new challenges for individual and collective privacy, security, as well as democracy and participation. The main objective of the CSA e-SIDES is to complement the research on privacy-preserving big data technologies, by analysing, mapping and clearly identifying the main societal and ethical challenges emerging from the adoption of big data technologies, conforming to the principles of responsible research and innovation; setting up and organizing a sustainable dialogue between industry, research and social actors, as well as networking with the main Research and Innovation Actions and Large Scale Pilots and other framework programme projects interested in these issues. It will investigate stakeholders' concerns, and collect their input, framing these results in a clear conceptual framework showing the potential trade-offs between conflicting needs and providing a basis to validate privacy-preserving technologies. It will prepare and widely disseminate community shared conclusions and recommendations highlighting the best way to ultimately build confidence of citizens and businesses towards big data and the data economy.

e-SIDES is based on the assumption that improving the dialogue between data subjects and big data communities will improve confidence in big data technologies. The confidence of citizens is only justified when the relevant actors and stakeholders have appropriate consideration and attention for ethical and societal issues of big data technologies. This sums up the two main goals of e-SIDES, i.e., to realise appropriate consideration and attention for ethical and societal issues in big data technologies and to improve the dialogue between data subjects and big data communities and, thereby, to improve the confidence of citizens towards big data technologies and data markets.

2.2. Structure of the deliverable and relationship to other deliverables

The present deliverable serves as an initial research framework, laying foundations for the further activities within the project. The idea behind such framework is to obtain the common ground regarding the key concepts to be applied, the scope of the project and its main goals.

With this idea in mind, the present document is divided into sections, corresponding to different goals pursued in the project and its focus areas. First, the general part describes the scope and objectives of the entire e-SIDES project and explains the key terminology to be applied throughout the project in order, on the one hand, to delineate the relevant subject matter of the undertaken research and, on the other, to ensure the coherence between different sections of the project. Second, the document explains the methodology and strategy for building the community of stakeholders who through an ongoing dialog will play a key role in informing the research activities undertaken within the project, enabling the 'big data on the ground' vision of this phenomenon. Third, the deliverable briefly explains the scope and focus of each substantive segment of the research agenda: the ethical, legal, societal and economic aspects of the big data technologies, including the related initiatives ongoing in the EU in these fields. Each part also includes the section containing relevant concepts to be used throughout the project.

Thus, the refined research framework serves as a first conceptual step explaining the overarching focus of the e-SIDES project in general and of different research segments in particular. Moreover, it serves the role of a common denominator in the sense of terminology and objectives to be achieved in the subsequent deliverables. Building on the shared goals and approach to the research agenda established in the Refined Research Framework, each partner of the consortium will subsequently pursue her respective research activities contained in the forthcoming deliverables. Updates of the content of this document will be provided via the e-SIDES website throughout the duration of the project in order to more accurately reflect the specific research choices and approaches adopted in the subsequent deliverables. Such updates include, for instance, further refinement of the terminology and further expansion of the glossary.

3. Conceptual framework and definitions

3.1. Scope

The enormous quantities of data that are collected, stored and processed may considerably improve the speed, effectiveness and quality of decision-making and innovation for companies, governments and researchers. The potential of Big Data and Analytics (BDA) is proving adept in the discovery of novel trends, patterns and relationships. As noted by the OECD¹, the social and economic value of data is mainly reaped when data are transformed into information and knowledge (gaining insights) and then used for decision making (taking action): this is enabled by the ubiquitous “datafication” of the physical world and by the new pervasive power of data analytics, as shown in the data value cycle shown in Figure 1.

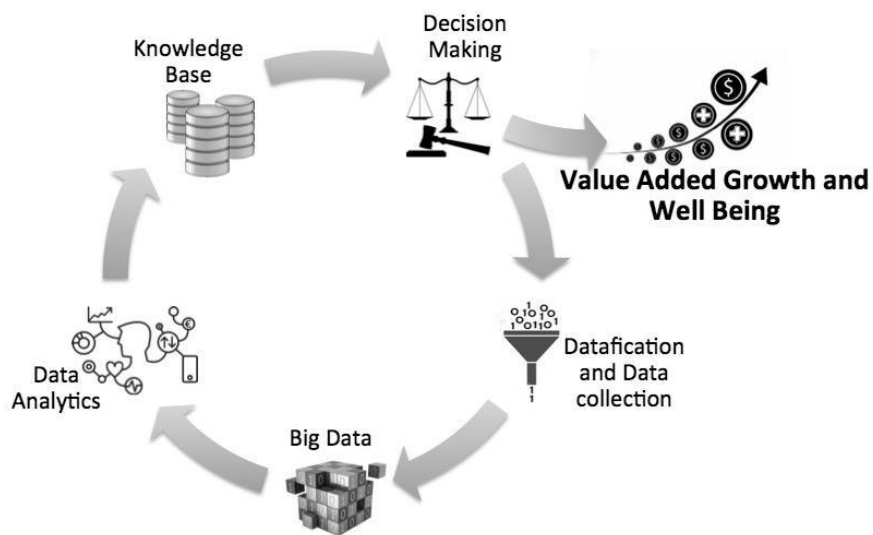


Figure 1 The data value cycle, Source: Data Driven Innovation OECD Oct. 2015

¹ OECD, *Data Driven Innovation for Growth and Well-Being*, October 2015

Data-driven innovation (DDI) is a catalyst for growth, competitiveness and productivity gains and is an essential ingredient of the digital transformation process occurring in all sectors of the economy, including low-tech industries and manufacturing. For example, according to a study based on 500 firms in the UK,² businesses exploiting online customer and strategic data (“datavores”) in their decision making process are 8-13% more productive than those that do not do so. Available evidence also shows that firms using DDI have raised productivity faster than non-users by approximately 5-10%.³ In a series of case studies on the impact of BDA, IDC has found evidence of higher productivity improvements (in the 15-20% range) and of economic benefits such as increased revenues (especially in customer and sales-related activities), for example in the retail, transport services, agriculture industries; and benefits of cost reduction and enhanced operational efficiency in production and operations, for example in manufacturing, banking and utilities. In the public sector, healthcare, research and education, the adoption of big data technologies and the diffusion of open data are leading to improvements in decision-making and knowledge processes, new and better services for citizens, and great benefits in transparency, openness and accountability. Overall, IDC has estimated that the European data economy could represent by the year 2020 approximately between 3% and 4.7% of the European Union (EU) Gross Domestic Product, corresponding to about €566-886 billion of direct and indirect impacts. However, this outcome is by no means certain and will require the development of a balanced demand-supply ecosystem, involving European SMEs in data-driven innovation, continuing R&D investments, creating favourable regulatory conditions for data access, re-use and control. But perhaps the most critical challenge will be to build confidence in the data economy, meeting the emerging privacy, security and power balance challenges emerging in a completely different social landscape characterised by massive information availability about everything and everybody.

Emerging social challenges include the potential loss of autonomy and freedom because of mass surveillance and discrimination enabled by data analytics; the risk of new types of market concentration and dominance in data value chains, due to “winner takes all” phenomena (think of the dominance of platforms such as Booking.com); risk of greater information inequalities leading to market power imbalances (between organizations, between citizens and governments, between consumers and suppliers).

Processing vast amounts of data may lead to situations in which data controllers may know many of the characteristics, behaviours and whereabouts of people.⁴ In some cases, data controllers may even know more about people than these people know about themselves, such as in cases of the life expectancies and happiness levels. Also they may be able to predict characteristics of people which are highly sensitive, including criminal records, sexual preferences and substance abuse. This may raise privacy issues, but also (other) ethical and societal issues, such as issues regarding discrimination, human dignity, justice, fairness and trust. In view of such privacy (and other ethical, legal and societal)

² Bakhshi, H., A. Bravo-Biosca, and J. Mateos-Garcia, (2014), *Inside the datavores: Estimating the effect of data and online analytics on firm performance*, Nesta, March, www.nesta.org.uk/sites/default/files/inside_the_datavores_technical_report.pdf

³ OECD, *Data Driven Innovation for Growth and Well-Being*, October 2015

⁴ Custers, B.H.M., Calders, T., Schermer, B., and Zarsky, T. (eds.) (2013) *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, Heidelberg: Springer.

considerations, businesses and governments are often unsure about how to deal with the data collected through their operations. On the one hand, the data is of particularly high value to companies and governments, enabling offering of personalised services or developing new business models, but, on the other hand, ethical and societal issues may result in lack of confidence, undermining efficient and legitimate data sharing and value creation for agreed purposes.

The challenge is, therefore, to develop technologies that are inherently privacy-preserving (or prevent discrimination, ensure human dignity, justice and fairness, increase trust and social acceptance, etc.). Within the research and innovation projects of call ICT-18-2016, such privacy-preserving big data technologies will be developed. Projects of call ICT-14-2016-2017 are likely to provide some insights on the real-world challenges and data to validate the privacy-preserving technologies.

e-SIDES is a Coordination and Support Action (CSA) that will complement the Research and Innovation Actions (RIAs) of the ICT-18 call on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies. e-SIDES will liaise with a broad and multidisciplinary community of stakeholders (including public administrations, research community, companies, civil society) to advise the research and innovation on privacy preserving big data technologies, promoting an integrated societally and ethically valid approach. Furthermore, e-SIDES will observe, map and report on Responsible Research and Innovation (RRI) issues in the field of big data, including technology, research, markets and education. e-SIDES will promote networking, awareness-raising and consultation among its communities, connect with technical RIAs to inform their thinking and issue reports, analyses and recommendations.

The focus of e-SIDES is on ethical and societal issues. When addressing ethical issues, legal issues are also taken into account, as these are typically closely related. For instance, some ethical principles and values are codified in EU legislation, which may provide more concrete requirements for the design of privacy-preserving big data technologies. Also, some ethical issues that are identified and validated in e-SIDES may feed recommendations on changing existing EU legislation. However, the scope of legal issues is restricted to the ethical issues of big data technologies. Legal issues that are not (also) ethical issues, such as difficulties regarding the enforcement of particular legislation, are beyond the scope of e-SIDES. All ethical and legal issues will primarily be mapped from a European perspective, focusing on the EU legal framework.

Whereas ethical issues focus on whether something is right or wrong, societal issues focus on how society is affected by something. Ethical and societal issues are usually related to each other. There are only a few examples that can be fully put under only one of the two categories. When addressing societal issues, economic issues are also taken into account, as big data technologies implicate economic issues that may have positive and negative societal consequences. Examples for such issues in the big data context are productivity growth and downsizing of work forces. Further issues that are taken into account from a societal perspective – and probably also from an ethical perspective – include trust, discrimination, inequality of access, exploitation and manipulation. With respect to societal and economic issues, e-SIDES focuses on the European multicultural and multinational landscape.

3.2. Goals

The broader, general goal of the project is twofold:

- To involve the variety of communities and stakeholders implicated in big data technologies to reach a common vision for an ethically sound approach to big data processing, and;
- To improve the dialogue between data subjects and big data communities (industry, research, regulators, civil society) and, thereby, to improve the confidence of citizens towards big data technologies and data markets.

It will be realized by focusing on a series of more specific objectives achieved throughout the project and connected to its different phases:

Objective I. Identifying and validating ethical and societal implications of privacy-preserving big data technologies

Objective II. Examining, discussing and validating privacy-preserving big data technologies;

Objective III. Liaising with researchers, business leaders, policy makers and civil society through open-access forums and community events;

Objective IV. Providing an open-access Internet-based meeting place for discussion, learning and networking regarding privacy-preserving technologies;

Objective V. Providing ethical-legal and societal-economic advice to facilitate responsible research and innovation in the field of big data technologies;

Objective VI. Observing, mapping and reporting on ethical and societal issues related to big data technologies, research, markets and education;

Objective VII. Providing an agreed-upon and collective community position paper presenting recommendations for each of the issues addressed in the community events to foster societally compatible and ethically valid big data research and innovation.

3.3. Terminology

Throughout the duration of the e-SIDES project, it is important to have a clear understanding of several specific terms in the domain of big data technologies and ethical, legal, societal and economic issues of these big data technologies. A clear terminology will help mutual understanding and create further focus in the e-SIDES project by including and excluding certain aspects and issues.

In this section, the most important (basic) terminology in the area of big data and big data technologies is provided. As such, this section focuses on technological terminology. The basic terminology regarding ethical, legal, societal and economic issues is provided in Sections 4.3, 5.3, 6.3 and 7.3, respectively. Furthermore, a glossary with brief descriptions of terms is provided in Section 9.

As the e-SIDES project progresses, the terminology described in this document will be further refined and the glossary will be further expanded. Updates of the terminology and the glossary will be provided via the e-SIDES project website.

When discussing technological terminology, perhaps the most difficult term to define or describe is **big data**. The term big data has become a buzzword and has become so popular that it is used in many different meanings. In some contexts, people use the term big data to describe datasets they consider

big. This is a rather subjective perspective, however. Datasets with thousands of records may be big for someone who is used to dealing with datasets with hundreds of records, but may seem small for someone who is used to dealing with datasets with millions of records.

There is no well-established definition of big data, but obviously its most prominent characteristic is its sheer volume. That is why it is called big. There is no generally accepted definition of the size required for datasets to be called big data. Big data can involve many terabytes or even petabytes of data.⁵ But it is not only size that matters. In general, big data refers to data and analytics that involve the 3 Vs: volume, velocity and variety.⁶ *Volume* refers to the large amounts of data and the fact that no samples are taken. *Velocity* refers to the fact that many data are real time or nearly real time. *Variety* refers to the different nature of various types of data, including text, numbers, images, videos and sound. Some authors add other Vs to these three, such as veracity (focusing on the inherent limited reliability of some data sources), variability (focusing on fluctuating sizes of data streams) and value (focusing on the limited added value of raw data in relation to its volume).

These characteristics of data pose significant challenges to analytics. Typically, big data does not allow the human eye or human intuition to grasp the essence of the data and the patterns and relationships hidden in it. Such hidden knowledge can usually only be discovered using advanced methods for data analysis.

Data is increasingly considered the new oil of the economy. Data has become a major raw material of production as well as a key source of economic and social value.⁷ In other words, data fuels the economy, more specifically, the data economy. It reveals completely new opportunities. For instance, data enables creating and delivering a new range of products and services, and accessing new groups of customers. It is expected that big data will accelerate these developments.⁸ Big data will become a key element of competition underpinning new waves of productivity growth, innovation and consumer surplus.⁹

All this sounds promising, but concrete examples of big data applications and the big data revolution are not abundant. Critics state that 'Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...'¹⁰ It is correct that not all large datasets qualify for big data applications and, thus, claims should be

⁵ Gandomi, A., & Haider, M. (2015) *Beyond the hype: Big data concepts, methods and analytics*, International Journal of Information Management, 35, p. 137-144.

⁶ Laney, D. (2001) 3D Data Management: Controlling Data Volume, Velocity and Variety. Gartner. Stamford CT: META Group Inc.

⁷ Tene, O. and Polonetsky (2012) *Privacy in the Age of Big Data; A Time for Big Decisions*, 64 Stanford Law Review Online, Feb 2, 2012, p. 63-69.

⁸ Mayer-Schönberger V. & Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*, New York: Houghton, Mifflin, Harcourt Publishing Company.

⁹ Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburg, C. and Hung Byers, A. (2011) *Big Data: The next frontier for innovation, competition and productivity*, Seoul: McKinsey Global Institute. See <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>

¹⁰ Woodcock, C. (2014) Big data is a lot like teenage sex, *Tech City News*, 7 August 2014. See <http://techcitynews.com/2014/08/07/big-data-is-a-lot-like-teenage-sex/>

evaluated critically. However, it is not true that no one knows how big data works. Large companies in all types of industries are dealing with big data and making or increasing their profit with it. Examples include retail organizations such as Walmart, Internet companies such as Google and Facebook, banks and credit card companies such as JP Morgan Chase and MasterCard, insurance companies such as AXA and Allianz, and navigation and mapping product producers such as TomTom and Garmin. Also an entire industry has grown around offering services for big data analytics, including companies such as Acxiom and Palantir.

In short, when looking at big data, it is defined in a way that new information technologies are required for storing, processing, analyzing, disseminating and managing the data. These new information technologies are termed as **big data technologies**. Some of these big data technologies focus on the storage of big data or data management. Perhaps the most important category of big data technologies, however, is that of **big data analytics**. These are the tools that ought to make some sense of large amounts of fast and unstructured data. In a way these technologies create value by transforming **data** into information useful for making decisions.¹¹

Now, some of these big data technologies or the way how they are applied may cause ethical and societal issues. These issues will be mapped extensively in the e-SIDES project, in Deliverable 2.2. Such issues involve privacy, human dignity, discrimination, security, trust and many other issues. These issues can be addressed in several different ways. One way is to create new technologies that address issues of previous technologies. A typical example of such a technological solution are anonymizers. They anonymize personal data with the aim to mitigate privacy issues. Such technological solutions are called **Privacy Enhancing Technologies (PETs)**.

Although such technological solutions may be helpful in addressing ethical and societal issues, the e-SIDES project does not focus on these types of technological solutions only, but rather also on a new approach, in which new technologies are designed in ways that they actually are already privacy preserving when they are ready for their first time use. The idea is to distil design requirements from ethical, legal, societal and economic perspectives and to help building new technologies that take the key requirements into account. In other words, privacy requirements are taken into account throughout the entire engineering process. This is what is called **Privacy by Design**.

Obviously in the term privacy by design, can be transformed into other equivalents, such as ethics by design, legitimacy by design, trust by design and equality by design. To avoid a plethora of terminology, we will use the term privacy by design as a term in which technologies are designed in such a way that they take into account all kinds of ethical, legal and societal issues.

A typical example of a privacy by design approach is so-called **privacy-preserving data mining**. When processing big data, **data mining** is a set of analysis tools that performs automatic analysis of data using mathematical algorithms, in order to find new patterns and relations in data.¹² These tools can be

¹¹ Custers B.H.M. (2013), *Data Dilemmas in the Information Society*. In: Custers B.H.M., Calders T., Schermer B., Zarsky T. (red.) *Discrimination and Privacy in the Information Society*. nr. 3. Heidelberg: Springer. 3-26.

¹² Calders T. & Custers B.H.M. (2013), *What is data mining and how does it work?*. In: Custers B.H.M., Calders T., Schermer B., Zarsky T. (red.) *Discrimination and Privacy in the Information Society*. nr. 3. Heidelberg: Springer.

designed in way in which such data mining preserves privacy. Hence, privacy-preserving data mining refers to data mining technologies with built-in requirements (based on privacy by design) to preserve privacy.¹³

4. Community building, stakeholders engagement

This chapter presents an overview of the approach designed and implemented by e-SIDES to build a community of stakeholders in the domain of big data and privacy-preserving technologies that will be involved in the various activities of the project. The detailed plan of stakeholder engagement and community building is presented in the Deliverable D6.1.

As a Coordination and Support Action (CSA), e-SIDES' impact will principally be achieved through externalizing or disseminating the results and knowledge generated through the series of community events that are planned. Indeed, among the expected impacts of the project we find:

- To ensure the effective involvement of the relevant stakeholders;
- To improve the dialogue between citizens/ individuals and big data communities (industry, research, policy makers, regulators).

Therefore, the stakeholder engagement is key to the success of the e-SIDES project and will be implemented through a multi-channel approach. e-SIDES will foster the stakeholder dialogue through offline and online activities: events, an online community platform and social media engagement. Besides these activities, e-SIDES will produce a set of community papers collecting the stakeholders' views around the issues analyzed. To ensure a balanced representativeness of different viewpoints, the papers will be published online on the project website to collect comments from the community.

The stakeholder dialogue established through e-SIDES events and through the online channels will serve as foundation for the entire project. Building a continuous interaction with the key target groups is essential to achieve the collaborative production of the community papers which, as the final output of the project, will collect the views of the different audiences and identify the priorities around the topics researched by e-SIDES.

To give home to the stakeholder community, e-SIDES has developed an online platform that will be the main hub of the online interaction and engagement. The platform will be hosted on the project's website www.e-sides.eu.

To perform an effective stakeholder engagement, e-SIDES will focus on the following activities:

- Stakeholder identification

¹³ Custers B.H.M., Calders T., Schermer B. & Zarsky T.Z. (2013), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, Studies in applied philosophy, epistemology and rational ethics nr. 3. Heidelberg: Springer.

- Awareness raising
- Community building and engagement

4.1. Stakeholder identification

The first step of the engagement process is the identification and selection of the appropriate audience. The selection process has encompassed the following phases:

- Identification of any relevant stakeholder group;
- Mapping of the groups' relationships to the project objectives and their relevance;
- Selection of the most relevant groups.

In the first part of the project, e-SIDES performed this mapping exercise, which led to the identification of 4 groups and relevant profiles and roles as key target audience:

- Big data research and academic communities
 - H2020 Research and Innovation Actions (RIAs) and Large Scale Pilot (LSP)
 - Universities
 - Researchers
 - Research associations
- Industry representatives
 - SMEs & start-ups
 - Medium/large companies
 - Data owners/generators
 - Industry associations
- Public sector
 - European Commission
 - Data protection authorities
 - Governmental agencies
 - Policy makers
- Civil society
 - Associations representing citizens
 - NGOs

The identification of relevant stakeholders is an ongoing process, and the whole consortium will contribute to this task. To ensure a good mix of stakeholders, e-SIDES will pay attention to represent in a balanced way the different categories, considering geography, gender and other relevant socio-demographic characteristics of the European landscape. e-SIDES will evaluate each time the most appropriate opportunity, channel and message to involve the different target groups to ensure effectiveness of the engagement and maximize the outputs of the project.

4.2. Awareness raising

During the first phase of the project, e-SIDES aims at generating visibility among the relevant audience groups identified and presented above, focusing on the general presentation of the project, overview of its objectives and expected results.

Progressively, in the advanced phase of the project, e-SIDES' effort will be focused on disseminating the achieved results and promoting the conclusions to the target community. As final output, e-SIDES will produce a set of community papers that will present the view of the big data community on privacy-related issues and how they may be encountered. The production of the output builds on an ongoing online and offline stakeholder dialogue, for validation and consensus building.

To ensure that the project reaches the target audience, e-SIDES will apply a multi-channel approach spreading communications via:

- The project website/online platform: www.e-sides.eu
- Workshops: the list of events will be uploaded on the website, shared on the social media, and periodically updated
- Communication materials: leaflets, brochure
- Position papers: e-SIDES will produce and publish online 7 position papers collecting the main outcomes of the stakeholders dialogue
- Social media: the project's social media channels are presented in the D6.1 deliverable

In terms of awareness raising, e-SIDES aims at achieving the following goals:

- Attracting potential collaborators and contributors to the draft community papers
- Attracting policy makers to provide insights and contribute to the identification of priorities
- Attracting privacy-preservation stakeholders from industry and civil society in the project events and on the online community

4.2.1. Community building and engagement

The community participation will be leveraged to investigate the main ethical and societal issues, assess their relevance for the different groups, discuss the willingness to accept certain technologies and the main factors affecting confidence in data technologies.

To facilitate the involvement and wide collaboration of the community, e-SIDES will highlight the opportunity to have access to information and insights and to influence the development of forthcoming technologies, as well as to shape the policy agenda at the EU and national level. Moreover, e-SIDES will offer concrete advice on ethical, legal, societal and economic issues to actors and stakeholders. The community will be involved and will have the opportunity to provide their contribution to the assessment of privacy-preserving big data technologies under development, or newly developed, on the extent to which they address the identified issues. The assessment will be particularly of interest for the industry stakeholders as it will lead to the identification of key design requirements and to the development of recommendations.

e-SIDES will involve actors and stakeholders periodically (rather than intermittently) over the entire duration of the project. e-SIDES has designed a set of complementary activities to reach out to the target audience:

- Organization and participation in events
- Creation and management of an online platform
- Interaction on social media

e-SIDES will organize 7 community events and a final conference. For each event, a discussion paper will be prepared as foundation for the discussion. The main outcome of each event will be a white paper summarizing the main results. It will be published on the online platform to collect contributions and validation from a broader community, beyond the events' attendees.

The online platform will be the home to the online community, and will facilitate interaction with the users and access to the content. The online discussion will be moderated by the community manager supported by the domain experts of the consortium.

A consistent part of the community interaction will be dedicated to leveraging synergies with other EC funded projects. RIAs and LSPs will be informed constantly about the content updates on the online portal, invited to the events and involved in the finalization of position papers.

In the advanced phases of the project, e-SIDES will focus on the collaborative development of the final set of recommendations by leveraging contributions of the whole community: each recommendation will be discussed at the events, and will also be published on the online portal for comments and revisions. The final aim is to build consensus around a core set of common principles and recommendations. The results will be published in the final community position paper and ideally signed by a representative sample and opinion leaders from the identified groups of stakeholders, including most of the relevant RIAs and LSPs.

5. The ethical perspective

5.1. Introduction

Big data technologies carry enormous promises for many different fields of human existence: from the area of health care to crime prevention to wealth redistribution – data analysis can help discovering methods for increasing overall societal well-being. However, “[a]s is often the case with the cutting edge of scientific and technological progress, understanding of the ethical implications of Big Data lags behind.”¹⁴ The ethical perspective on the big data technologies could be perceived as an endeavour in understanding and presenting which aspects of human life as an individual as well as a society are impacted by this new technological realm through the prism of the foundational values indicative for which actions are considered to be moral. While such values, as human dignity, autonomy, liberty, equality, are also the basis for legal rights, this part of the project will look at them *in abstracto*, as guiding ethical framework for conceptualization of the existing and also emerging big data practices. Thus, for the purposes of this section, it is essential to determinate the key transformative features of big data technology and scrutinize them in the context of the values which are considered relevant for the development of an ethically sound informational and technological architecture in Europe. The perceived gains and promises brought about by data technologies, or so called ‘data fundamentalism’, “problematically influence[s] the public, mass media and researchers where a tendency exists to view the advancement of Big Data into all information-based disciplines as inevitable.”¹⁵ Specifically, however, the e-SIDES project takes the view that certain attitudes and conceptions towards technological developments should not be perceived as fixed, following the path-dependent tracks, but rather should undergo constant re-evaluation pursuant to the ethical guiding map.

Consequently, this part of the project concerned with the ethical perspective on the big data technologies aims at:

- a) developing an overview of the most relevant ethical principles guiding the normative choices in the realm of information and data technology;
- b) establishing which transformative features of big data technologies might challenge such guiding ethical framework and in which manner;
- c) proposing conceptual approaches towards the development of big data technologies which might contribute to safeguarding ethically sound technological progress.

It is important to distinguish the ethical perspective from the legal perspective (discussed in the next section). The ethical perspective focuses on values that may need protection (such as human dignity, autonomy, liberty and equality), whereas the legal perspective, particularly the fundamental rights perspective, focuses on how these values can be protected with legal instruments (such as the right to privacy, the right to equality, etc.).¹⁶ In other words, human rights try to protect particular values (the

¹⁴ Mittelstadt B., Floridi L. (2015), *The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts*, Science and Engineering Ethics, published online 23 May 2015.

¹⁵ Ibidem

¹⁶ Note that these values can also be protected in other, non-legal ways, for instance by organisational measures (like security measures) or by technological measures (like privacy by design).

legal perspective), whereas ethics focuses on the object that requires protection (the ethical perspective).

Furthermore, it is important to distinguish the ethical perspective from the societal perspective (discussed in section 6). The ethical perspective focuses on individual human beings (including their vulnerabilities, needs, interests and preferences), whereas the societal perspective focuses on groups of people. As such, the societal perspective is, in a sense, an aggregated perspective, which may be different from the individual perspective. A typical example to explain the difference is the paying of taxes: nobody likes to pay taxes and from an individual perspective not many people would voluntarily pay taxes when asked so. However, from a societal perspective, it makes sense to pay taxes, as it allows society to enable things (like infrastructure, police, a health care system) that individuals are unlikely to create themselves.

5.2. Focus and related initiatives

The focus of this part of the e-SIDES project can be described along the lines of the ethical, technological and methodological frameworks implicated.

First, the ethical framework considered in this part of the project will be informed by the current fundamental rights framework accepted in the EU. Thus, the focus will be on distilling those values which are foundational to the human rights instruments adopted in the EU. This approach will enable building upon the ethical approach to data technologies' development in further stages of the project and coherence throughout the analysis in its different parts.

Second, the project will focus on these aspects of big data technologies which carry certain transformative aspects toward the *status quo* of the current informational realm. The idea behind the analysis is to establish how such new aspects of big data technologies challenge the accepted ethical principles and how the latter could be conceptualized to impact the technological development in a meaningful way.

Thirdly, the methodological focus will be on reviewing the current literature on the ethical aspects in different fields of big data technologies and establishing certain common threads regarding their transformative impacts and resulting challenges from the ethical perspective. Furthermore, the crucial part of the project will be consulting the common perceptions and attitudes of different stakeholders towards the identified ethical challenges of big data technologies.

The ethical approach to big data technologies part of the e-SIDES project will build upon the findings presented in different previous or ongoing initiatives, including:

- CANVAS: Constructing an Alliance for Value-driven Cybersecurity¹⁷ (started in 2016) - aiming to unify technology developers with legal and ethical scholar and social scientists to approach the challenge how cybersecurity can be aligned with European values and fundamental rights;

¹⁷ <https://canvas-project.eu>

- RESPECT¹⁸ (2012–2015) - drawing up professional and ethical guidelines for the conduct of socio-economic research. Among other aspects, it focused on research ethics and data protection;
- SAPIENT: Privacy, Ethics and Rights as Standards for European Surveillance practices¹⁹ (2011–2014) - specifying for policy makers, technology developers and other stakeholders how and when smart surveillance should be used and the criteria to verify that surveillance systems respect the privacy of citizens;
- PRESCIENT: Privacy and emerging fields of science and technology: ethical, social and legal aspects²⁰ (2010–2013) - identifying and assessing privacy issues posed by emerging sciences and technologies and contributed to the development of new instruments for the governance of science and technology.

5.3. Terminology

Within the ethical perspective it is important to distinguish between moral values (or ideals), moral principles and moral rules. Ideals and values are more abstract and general than principles, which, in turn are more abstract and general than rules. Ideals or values can be considered as the desirable state of affairs that one aims to realize. Principles are the norms prescribing how to achieve this state of affairs. Rules (moral rules, which sometimes may also be legal rules) prescribe norms of conduct, e.g., do not steal, do not smoke, pay your bills, etc. Principles prescribe the approach to realize ideals or values to the highest degree possible. Principles can be fulfilled to a certain degree, whereas rules are either fulfilled or not.

Consequentialism

A normative theory in moral philosophy whereby the undertaken choices are judged pursuant to the results they bring about rather than the means for achieving these results. According to the consequentialist theories, the morally acceptable choices are these which maximize certain outcome that is perceived as intrinsically valuable.

Contextual integrity

The term coined by H. Nissenbaum,²¹ conceptualizing privacy as a state where flows of information conform to the expectations of the individuals about which parts of their identity are revealed in specific contexts. Given the fact that such expectations may change not only over time but also from one context to another, the notion of contextual integrity perceives privacy as a dynamic concept that should enable to control the flows of information pursuant to different variables of time, place and interpersonal relationships, etc.

Deontology

¹⁸ <http://www.respectproject.org>

¹⁹ <http://www.sapientproject.eu>

²⁰ <http://www.prescient-project.eu>

²¹ Nissenbaum, H. (2010) *Privacy in Context*, SUP.

Contrary to consequentialist theories, the deontological group of views is based on the moral evaluation of the action itself, rather than the result of such action.

Utilitarianism

A normative doctrine of consequential character pursuant to which the morally just actions are these which promote the greatest happiness of majority.

Virtue ethics

A normative theory in moral philosophy which has the virtues of character of an individual at its core (just like consequentialists have the outcome of the actions and deontologists the moral duties connected to the action as such).

6. The legal perspective

6.1. Introduction

Just like the ethical perspective on the data technologies development in Europe focuses on the values which are at stake in such technological contexts, the legal perspective takes as its point of departure the legal rights, enshrined in binding legislation, aimed at protecting these values. The departing point for analyzing the challenges of big data technologies from the legal perspective is outlining the core European normative commitments, stemming from the shared values and encapsulated in the set of fundamental rights, which might play a role in shaping emerging data technologies or are perhaps already at stake in different technological domains where data is gathered and processed on a massive scale. Pursuant to Article 2 of the Treaty on European Union “[t]he Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail”.²² The e-SIDES project is premised on the view that while potentially of great value for technical progress in various domains, big data technologies bring about also certain challenges to these core normative commitments. The idea behind the project is thus to delineate the ‘European way’ towards big data: to map out which fundamental rights are of relevance in the technological environments specific for big data²³ and see how these fundamental rights could be taken into account in order to ensure that already existing technologies and the ones being developed conform with the values relevant for the European society.

This part of the project will not take the positivistic approach to law by only looking into the binding legislation and its application to the technological *status quo*, but rather will approach the fundamental rights framework as a set of rights which enter into constant dialogue with different forces that shape the human condition at any given time (like existing social norms or technology) and thus should be interpreted in light of such changes. Hence, the idea behind the approach taken in this project is that scope of fundamental rights should be treated as a given, but rather construed in response to such evolving context.

Consequently, in this section e-SIDES aims at:

²² Treaty on European Union, consolidated version signed on 13.12.2007, OJ C 326 , 26/10/2012 p.1 – 390.

²³ Pursuant to the White Paper of the Council for Big Data, Ethics and Society, “the bigness of big data points to the newly expansive capabilities to connect disparate datasets through algorithmic analysis, forging unpredictable relationships between data collected at different times and places and for different purposes. Although it is impossible to identify a single characteristic that makes data ‘big,’ the emergent properties of massive, connected, and heterogeneous datasets are different than those of ‘traditional’ datasets that remain restricted to a context much closer to their original point of collection”. Metcalf, Jacob, Emily F. Keller, and danah boyd, 2017, *Perspectives on Big Data, Ethics, and Society*, Council for Big Data, Ethics, and Society. Accessed March 20, 2017, available at: <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/>.

- a) Setting forth and analyzing the fundamental rights framework as currently applicable in Europe which is of relevance for big data technologies;
- b) Looking at the development of case law and legislative initiatives in applicable fundamental rights domains in function of the challenges of the new technologies, taking into account the temporal perspective and main drivers that shaped the current EU fundamental rights framework in the context of new technologies (in order not to approach certain legal solutions as a given and avoid path-dependency in this aspect, but rather look at how they were developed overtime and thus ultimately how could be further adapted towards changing technology);
- c) Scrutinizing how the fundamental rights framework currently in place informs big data technologies and which new initiatives could be undertaken in order to better align such technologies with the values stemming from this framework (taking into account different strategies of such alignment, including new legislative measures aimed at different applications of new technologies, requirements of privacy by design or tools or measures based on different cognitive processes, such as ‘nudges’).

6.2. Focus and related initiatives

The focus of the e-SIDES project at the nexus of big data technologies and law can be described along different lines taking as its main factor the jurisdictional aspects, substantive aspects and methodological aspects of the project.

First, with regard to the jurisdictional aspects, the project will take the ‘European approach’. This means that it focuses on the fundamental rights legislation binding in Europe at the supranational level, in particular within the EU and Council of Europe. Consequently, the main legal sources to be taken into account are the Charter of Fundamental Rights of the European Union (EU Charter)²⁴ and the European Convention on Human Rights (ECHR).²⁵ Additionally, the EU secondary legislation on the issues relevant to the analyzed fundamental rights will be discussed. Moreover, the project will closely look at the case law of the Court of Justice of the European Union (CJEU) as well as the European Court of Human Rights (ECtHR) pertaining to this domain.

Second, on the substantive level, the project will look at different fundamental rights relevant in the context of big data technologies. By analyzing the new features brought about in different social contexts by big data technologies, the research will map the main vulnerabilities of natural persons in such context and connect them to the fundamental rights which could play a role in safeguarding the values at stake. The list of the fundamental rights triggered in the context of big data would include not only the rights to privacy and data protection, where Europe is at the forefront of development of the legal protection²⁶, but also different issues connected to the right to equality, non-discrimination, freedom of expression, dignity. Based on the academic literature as well as case law and the ethical

²⁴ European Union, Charter of Fundamental Rights of the European Union [2000] OJ C364/01 and [2010] OJ C83/389.

²⁵ European Convention on Human Rights and Fundamental Freedoms, 1953.

²⁶ See for example: Pollicino, O., Kowalik-Bańczyk, K. (2016) *Migration of European judicial ideas concerning jurisdiction over Google on withdrawal of information*, German Law Journal, Vol. 17 No. 03, 2016.

section of the e-SIDES project, we will distil the main values underpinning the identified fundamental rights, see how these values are challenged by different applications of big data technologies and how they can inform the regulatory initiatives undertaken in this field. We will also undertake the temporal analysis of the fundamental rights in question, so as not to approach them in a static manner, as a given and completely entrenched in the European legal landscape, but rather to trace the evolution of the meaning and the scope of different rights where they enter the realm of new technologies.

Finally, from the methodological perspective, the focus of the project in this section, as in the other ones, will be on interactions with different stakeholders, whose input will inform the analyzes. Thus, the legal research will be connected to the most acute issues relating to the big data technologies within different communities, as identified during the ongoing dialog undertaken with these communities as a part of e-SIDES. Consequently, the theoretical findings will be paired with the empirical ones giving the up to date outlook on the legal challenges of the big data.

- e-SIDES draws on the findings of a series of previous and ongoing research projects, both EU-funded as well as undertaken on a national or international level. The projects relevant for the present section include:
- CANVAS: Constructing an Alliance for Value-driven Cybersecurity²⁷ (started in 2016) - aiming to unify technology developers with legal and ethical scholar and social scientists to approach the challenge how cybersecurity can be aligned with European values and fundamental rights;
- RESPECT²⁸ (2012–2015) - drawing up professional and ethical guidelines for the conduct of socio-economic research. Among other aspects, it focused on research ethics and data protection;
- PACT: Public perception of security and privacy²⁹ (2012–2015) - assessing existing knowledge about the relation between security and privacy and the role played by trust and concern;
- SurPRISE³⁰ (2012-2015) - assessing of criteria and factors determining acceptability and acceptance of security technologies in Europe, re-examining the relationship between privacy, fundamental rights and security;
- SAPIENT: Privacy, Ethics and Rights as Standards for European Surveillance practices³¹ (2011–2014) - specifying for policy makers, technology developers and other stakeholders how and when smart surveillance should be used and the criteria to verify that surveillance systems respect the privacy of citizens;
- CONSENT: Consumer Sentiment Regarding Privacy on User Generated Content³² (2010-2013) - examining how consumer behaviour and commercial practices are changing the role of consent in the processing of personal data;
- PRESCIENT: Privacy and emerging fields of science and technology: ethical, social and legal aspects³³ (2010–2013) - identifying and assessing privacy issues posed by emerging sciences and

²⁷ <https://canvas-project.eu>

²⁸ <http://www.respectproject.org>

²⁹ <http://www.projectpact.eu>

³⁰ <http://surprise.project.eu>

³¹ <http://www.sapientproject.eu>

³² <http://www.consent.law.muni.cz>

technologies and contributed to the development of new instruments for the governance of science and technology;

- Council for Big Data, Ethics and Society³⁴ - developing frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin the big data phenomenon.

6.3. Terminology

Court of Justice of the European Union (CJEU)

The judicial body of the EU, located in Luxembourg, whose tasks consist in ensuring the compliance with the EU law by member states, reviewing the legality of the acts of EU institutions and interpreting the EU law at the request of the national court and tribunals (the procedure of the reference for the preliminary ruling).

Data controller

Pursuant to Article 3(8) of the Data Protection Directive, it means the competent authority which, alone or jointly with others, determines the purposes and means of the processing of personal data.

Data subject

Pursuant to Article 3(1) of the Data Protection Directive, it means identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to 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 of that natural person.

Data protection

It can be understood as one of the fundamental rights, protected under Article 8 of the EU Charter, which is the first fundamental right instrument to specifically establish the right to data protection as a separate fundamental right, distinguished from the right to privacy. The right to data protection as such was included into the EU fundamental rights framework as a response to the challenges of the new technology with respect to data gathering and processing in both public and private spheres.³⁵ The relationship between the right to data protection and the right to privacy remains contestable, as the subject matter of these rights in the large part overlap and the rights themselves have been often conflated in the CJEU case law.³⁶ The specific scope of the right to data protection is delineated in the secondary legislation (currently the Data Protection Directive and soon the General Data Protection Regulation). Data protection as a term can however be also understood more broadly, as a legal

³³ <http://www.prescient-project.eu>

³⁴ <http://bdes.datasociety.net>

³⁵ EU Network of Independent Experts on Fundamental Rights, Commentary of the Charter of Fundamental Rights of the European Union, June 2006, p. 90, available at: ec.europa.eu/justice/fundamental-rights/files/networkcommentaryfinal_en.pdf

³⁶ Lynskey, O. (2015) The Foundations of EU Data Protection Law, OUP 2015.

framework aimed at ensuring control over flows of personal data to its subject and safeguarding against unlawful processing of such data (including different primary and secondary law legislation which together aim at protecting individuals' right to protect her data).

Data protection directive

The Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. This is the EU secondary law, binding on the EU member states indirectly, to be transposed into national laws by the end of 1998. The directive has a double purpose of ensuring the protection of the fundamental right of data subjects as regards processing their personal data (the fundamental right safeguarding purpose) and also enabling free flow of this data between the EU member states (the common market enabling purpose).³⁷ It is the first EU law legislation on data protection and on 24 of May 2018 it will be replaced by the General Data protection Regulation. The Directive outlines the scope of the right to data protection and sets forth the principles for legitimate processing of data.

EU Charter of Fundamental Right (EU Charter)

The Charter of Fundamental Rights of the EU was first proclaimed in 2000 and subsequently entered into force in December 2009 with the entry into force of the Treaty of Lisbon. The legal value of the EU Charter is such as the value of the EU Treaties (Article 6.1 Treaty on the EU). The Charter brings together all the fundamental rights binding in the EU, including the 'new' fundamental rights, which constitute a response to the technological challenges in modern societies (on of such new fundamental rights being the right to data protection). The provisions of the Charter apply to the 'institutions, bodies, offices and agencies of the Union with due regard for the principle of subsidiarity and to the Member States only when they are implementing Union law' (Article 51.1 of the EU Charter). In case of overlap between the rights set forth in the EU Charter and those guaranteed under the ECHR, 'the meaning and scope of those rights shall be the same as those laid down by the said Convention' (Article 52.3 of the EU Charter).

European Convention on Human Rights (ECHR)

The Convention is an international treaty drafted in 1950 within the Council of Europe which entered into force in 1953 which stipulates the number of civil and political rights as previously enshrined in the Universal Declaration of Human Rights.³⁸ The ECHR was ratified by all the 47 Council of Europe member states, including all the EU member states. The rights included in the Convention are interpreted by the European Court of Human Rights and have been transposed by the member states at the national level.

European Court of Human Rights (ECtHR)

The ECtHR is an international human rights court based in Strasbourg, ruling on applications filed by individuals or member states (in case of inter-states disputes) concerning the breaches by the Council of

³⁷ Ibidem

³⁸ Universal Declaration of Human Rights, adopted by UN General Assembly Resolution 217 A(III) of 10 December 1948.

Europe member states of the fundamental rights stipulated in the ECHR. The rulings of the ECtHR are binding on the states.

Fundamental/human rights

A set of rights giving expression to the idea that every individual, simply on the basis of her status as a human being, 'has legitimate claims upon his or her own society for certain freedoms and benefits'.³⁹ Fundamental rights in the EU were first recognized by the CJEU as 'enshrined in the general principles of Community law'⁴⁰ and subsequently included in the EU Charter.

General Data Protection Regulation (GDPR)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, will enter into force on 24 May 2018 and thereby be directly applicable law in all the EU member states. The GDPR aims at providing for more uniform protection of personal data throughout the EU and strengthening the rights of the data subject.

Privacy

The notion of the right to privacy was coined for the first time in 1890 by Samuel Warren and Louis Brandeis, in a scholarly article, pursuant to which the right to privacy was equated with the right to be left alone.⁴¹ The exact value and interests underpinning the right to privacy are debatable and can be derived from various concepts, such as liberty, personal autonomy or dignity.⁴² In the EU, the right has been recognized as a fundamental right in Article 8 of the ECHR and Article 7 of the EU Charter as well as in various constitutional provisions on the national level. The right to privacy is not absolute, as most of the human rights, but rather, in cases of conflict, should be balanced against other rights and interests pursuant to the principle of proportionality. The ECtHR has been resolving such conflicts on the case by case basis and thus the exact scope of the right to privacy is not fixed but rather can be adjusted depending on the specific circumstances of a given case.

³⁹ Henkin, L. (1989) *The Universality of the Concept of Human Rights*, The ANNALS of the American Academy of Political and Social Science, Vol 506, Issue 1, 1989.

⁴⁰ *Stauder*, C-29/69.

⁴¹ Warren, S., Brandeis, L. (1890) *The Right to Privacy*, 4 Harv. L. Rev. 193, 1890.

⁴² See for example: Westin, A. (1967) *Privacy and Freedom*, New York : Atheneum, 1967; Gavison, R. (1980) *Privacy and the limits of law*, Yale Law Journal 89 : 421–71, 1980, Post, R. (1989) *The social foundations of privacy: Community and self in the common law. Tort*, 77 California Law Review : 957–1010, 1989.

7. The societal perspective

7.1. Introduction

With respect to the societal perspective, e-SIDES takes into account existing attempts to structure and assess societal impacts and builds on them as far as possible. Of particular importance is the work done by the International Association for Impact Assessment (IAIA). However, other initiatives as well as previous and ongoing research projects are also considered. Regarding initiatives and research projects, particular attention is paid to activities that deal with big data technologies and their applications, and are funded by the EU or its Member States.

7.2. Focus and related initiatives

Due to the fact that societal (or social) impacts are conceived as being anything linked to an intervention (e.g., a policy, programme, plan or project) that affects or concerns any impacted stakeholder group, almost anything can potentially be a societal impact so long as it is valued by or important to a specific group of people. The IAIA argues that this means that societal impact assessments (SIAs) cannot start with a generic checklist of potential impacts but must identify the impacts from an awareness of the intervention whose societal impacts are to be assessed.⁴³ For e-SIDES, this underlines the need to work intensively with stakeholders affected by or involved in the development of big data technologies when selecting the key impacts to focus on in the project. To make the selection process as efficient as possible, impacts already documented in the literature will be brought into the discourse with the stakeholders by the e-SIDES consortium.

A list of societal impacts provided by Vanclay⁴⁴ gives a board overview of what societal impacts are. Changes to one or more of the following are considered as societal impacts:

- the way of life characterised by how people live, work, play and interact with one another on a day-to-day basis;
- the culture characterised by shared beliefs, customs, values and language or dialect;
- the cohesion, stability, character, services and facilities of communities;
- the political systems characterised by the extent to which people are able to participate in decisions that affect their lives, the level of democratisation that is taking place, and the resources provided for this purpose;
- the environment of people characterised by the quality of the air and water, the availability and quality of the food, the level of hazard or risk, dust and noise; the adequacy of sanitation, physical safety, and access to and control over resources;
- health of people in the sense of a state of complete physical, mental, social and spiritual wellbeing and not merely the absence of disease or infirmity;

⁴³ Vanclay, F., Esteves, A.M., Aucamp, I. And Franks, D. M. (2015) *Social Impact Assessment: Guidance for assessing and managing the social impact of projects*, International Association for Impact Assessment, Fargo ND, 2015. https://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf.

⁴⁴ Vanclay, F. (2003) International Principles for Social Impact Assessment. *Impact Assessment & Project Appraisal* 21(1), 5-11, 2003.

- personal and property rights, particularly whether people are economically affected or experience personal disadvantage, which may include a violation of their civil liberties;
- fears and aspirations including perceptions about their safety, fears about the future of their community, and aspirations for their future and the future of their children.

Without doubt, some of the changes are more relevant in the context of big data technologies and their application than others. Not surprisingly, the list shows that the economic perspective is usually considered part of the societal perspective. Due to the high relevance of economic impacts in the context of technology innovation and to facilitate the investigation of possible trade-offs between economic issues and issues of the other perspectives, it was decided to view societal and economic impacts separately.

Since not everybody is affected in the same way by big data technologies and their application, the e-SIDES consortium will pay particular attention to the demarcation of a preliminary social area of influence. This includes gaining a good understanding of the communities likely to be affected by preparing a community profile that will include a thorough stakeholder analysis, a discussion of the socio-political setting, an assessment of the differing needs, interests, values and aspirations of the various subgroups of the affected communities, an assessment of their impact history, a discussion of trends happening in those communities and a discussion of the assets, strengths and weaknesses of the communities.

According to IAIA⁴⁵, understanding the local cultural context is essential for the success of a SIA as well as the intervention whose societal impacts are assessed. It is important to acknowledge that different societies have different cultural values, with different understandings about how things should be done. These differences can pose many challenges, particularly if the focus lies on a large and diverse geographic area such as the EU. Gaining a good understanding of local contexts or at least being aware of local differences is critical with respect to the effectiveness of e-SIDES's participation processes.

In addition to the direct societal impacts, the indirect, second and higher order impacts also need to be considered and analysed. Careful attention needs to be given to thinking what these may be, especially because they may at first not be obvious. A brainstorming process will be used by the e-SIDES consortium to identify and structure impacts, as well as to support thinking about impact pathways and cumulative impacts. As the analysis is a mixture of ex-ante and ex-post assessments of interventions, it does not necessarily comprise only collecting and analysing data but also requires a comparison with experiences elsewhere, having experts input into scenario analyses and the use of reasoned thinking.

After the impacts have been identified, it is then necessary to prioritise them. In e-SIDES, criteria for establishing significance will be determined and each impact rated on the basis of those criteria. There are several methods that can be used to do this. It has not yet been decided if a multi-criteria analysis or a risk assessment methodology will be used. Based on the stakeholder groups identified, several

⁴⁵ Vanclay, F., Ana Esteves, A.M., Aucamp, I. and Franks, D.M. (2015) *Social Impact Assessment: Guidance for assessing and managing the social impact of projects*, International Association for Impact Assessment, Fargo ND, 2015. https://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf.

significance assessments may be conducted. The assignment of significance will most likely be made in workshop environments with the stakeholders themselves making the assessment.

Another relevant source is the SIAhub for Social Impact Assessment practitioners⁴⁶. SIAhub is an Internet portal where the global SIA community of practice can network, access resources, share ideas and promote good practice. Another relevant institution is Data & Society⁴⁷, which is a US-based research institute focused on the social and cultural issues arising from data-centric technological development.

Frameworks proposed for conducting SIAs are largely similar to those used for assessing other impacts or for risk management.⁴⁸ We are thus confident to use a common process for assessing not only societal but also ethical, legal and economic impacts.

Apart from the work done by the institutions mentioned above, the following initiatives and research projects are considered particularly relevant from the societal perspective

- BYTE: The Big Data Roadmap and Cross-disciplinary Community for Addressing Societal Externalities⁴⁹ (2014-2017) - The BYTE project assisted European science and industry in capturing the positive and diminishing the negative externalities associated with big data. Work conducted within the scope of the project can be useful for e-SIDES particularly with respect to the discovery of impacts and stakeholders relevant in the context of big data.
- SAPIENT: Privacy, Ethics and Rights as Standards for European Surveillance practices⁵⁰ (2011–2014) – The SAPIENT project specified how and when smart surveillance should be used and the criteria to verify that surveillance systems respect the privacy of citizens for policy makers, technology developers and other stakeholders.
- DESSI: Decision support on security investment⁵¹ (2011-2013) - The DESSI project provided a process and a decision support system to end users of security investments. The system gives insight into the pros and cons of specific security investments. It contributes to a transparent and participatory decision making that accounts for context and multi-dimensionality of society. The results of the project can be useful for e-SIDES with respect to the discovery of societal impacts.
- ASSERT: Assessing security research: Tool and methodology to measure societal impact⁵² (2013-2015) - The ASSERT project compared different assessment methodologies and synthesize a

⁴⁶ <http://www.socialimpactassessment.com/>

⁴⁷ <https://datasociety.net/>

⁴⁸ Hossein, M., Renn, O., Vanclay, F. et al. (2013) *A framework for combining social impact assessment and risk assessment*, Environmental Impact Assessment Review, Vol. 43, No. 1, 2013, pp. 1-8. <http://www.sciencedirect.com/science/article/pii/S0195925513000541>. Wright, D., and Friedewald, M., *Integrating privacy and ethical impact assessment*, Science and public policy, Vol. 40, No. 6, 2013, pp. 755-766. Wright, D., Wadhwa, K., Lagazio, M. et al., *Integrating privacy impact assessment in risk management*, International Data Privacy Law, Vol. 4, No. 2, 2014, pp. 155-170.

⁴⁹ <http://new.byte-project.eu/>

⁵⁰ <http://www.sapientproject.eu>

⁵¹ <http://securitydecisions.org/>

⁵² <http://assert-project.eu/>

process for assessing the impact of security research from different perspectives to help technology developers to develop and offer products and services in line with society's actual needs. A key project result is the development of an online assessment tool for determining societal impact, which is expected to guide researchers in evaluating their research efforts and goals. e-SIDES can make use of the list of societal impacts that ASSERT has identified and their methodology for interactively assessing technologies with the stakeholders.

- Socialising Big Data⁵³ (2013-2014) – The Economic and Social Research Council (ESRC) funded project developed a social framework for big data. Big data is considered to create vulnerabilities in part because of the tendency to overlook the social lives of data-objects, which are neither natural nor technical phenomena, but enacted and sustained through multiple and selective social practices, and hence always and already limited and limiting.

7.3. Terminology

The definitions were taken from IAIA.⁵⁴

Community

A commonly-used, yet contested concept that can apply at different levels, although generally referring to a place-based grouping of people who are presumed to have some sense of shared identity, some shared interactions of everyday life, and some common social and political institutions. Although individuals experience some social impacts at a personal level, the general assumption in SIA is that people live, work and play in social groupings called communities, which are therefore a primary focus in SIA.

Community profile

A description of the communities likely to be affected by a planned intervention.

Culture

The material and non-material aspects of a way of life that are shaped and transmitted among members of a community or a larger society. Sometimes referred to as the shared beliefs, values, norms, behaviours, language, and material objects that are passed from one generation to the next.

Cumulative impact

The successive, incremental and combined impacts of one or more interventions on society, the economy or the environment. They can result from the aggregation and/or interaction of impacts within a social or environmental system and are defined from the perspective of the people or environment experiencing them.

Direct impact

⁵³ https://www2.warwick.ac.uk/fac/cross_fac/cim/research/archive/socialising-big-data/

⁵⁴ Vanclay, F., Esteves, A.M., Aucamp, I. And Franks, D. M. (2015) *Social Impact Assessment: Guidance for assessing and managing the social impact of projects*, International Association for Impact Assessment, Fargo ND, 2015. https://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf.

An impact which occurs as a direct result of an intervention, which may also be called primary impact or first order impact. In e-SIDES, it refers to societal changes and impacts caused directly by big data technology and its application.

Ex-ante assessment

Most impact assessments are conducted in advance. They make predictions about the likely impacts of an intervention, about something that has not yet happened.

Ex-post assessment

Are evaluations of the impacts of a particular intervention.

Indirect impact

The indirect social impacts that happen after the immediate first order impacts in the chain of impacts that arise from an intervention. Indirect impacts may also be called higher order impacts.

Impact

An economic, social, environmental and other consequence that can be reasonably foreseen and measured in advance if a proposed action is implemented.

Impact assessment

The process of identifying the future consequences of a current or proposed intervention.

Mindmap

A mindmap is a diagram that is a way of presenting ideas and thoughts. Mindmapping is a process of organising information and ideas.

Social area of influence

Refers to the area over which an intervention creates impacts. Because of the mobility of people and the extent of social impacts, the social area of influence is likely to be much larger in physical area than the physical area of influence.

Social impact

Something that is experienced or felt, in a perceptual or corporeal sense at the level of an individual, social unit (i.e., family, household, collectivity) or community/society.

Stakeholders

Include all individuals and groups who are affected by, or can affect, a given intervention. Stakeholders consist of individuals, interest groups, and organizations.

Stakeholder analysis

A deliberate process of identifying all stakeholders of an intervention (i.e. the individuals and groups that are likely to impact or be impacted by it) and understanding their concerns about the project and/or relationship with it.

8. The economic perspective

8.1. Introduction

As was already stated before, the economic perspective is usually considered part of the societal perspective. In order to facilitate the investigation of possible trade-offs between economic issues and issues of the ethical, legal and societal perspectives, it was decided to view societal and economic impacts separately. This implies, however, that there are several overlaps between the initiatives and research projects referenced in this section and the section on the societal perspective.

In the e-SIDES project two rather different objectives are pursued with respect to the economic perspective:

- a) For one thing, e-SIDES takes economical issues such as de facto monopolies and changes in employment practices caused by big data technologies and their applications into account in the same way as it considers ethical, legal and societal issues;
- b) Additionally, future privacy-preserving big data technologies will have to compete with other big data technologies on the market. Assessing the economic viability of big data technologies and applications requires a thorough understanding of relevant costs and benefits.

8.2. Focus and related initiatives

Relevant institutions that have more specifically focused on the economic impact of big data include the OECD⁵⁵, IDC⁵⁶, the World Economic Forum⁵⁷ and the McKinsey Global Institute⁵⁸. The OECD, for instance, dealt intensively with the economics of data. It stresses that more and better data may lead to new goods and services, optimized business processes, better organizational management and faster innovation but also that increasing returns to scale and scope combined with multi-sided markets and network effects, favour market concentration and dominance.⁵⁹ Moreover, the OECD addressed, for instance, the shift in power that exacerbates existing inequalities and potential structural change in labour markets. IDC, for instance, recently studied the level of citizens' reliance on data as well as the gap between demand and supply of data worker skills⁶⁰. It is striking, that the majority of publications on the economic impact of big data concentrate on the benefits and do not pay much attention to potential costs. This is yet another reason to look into the costs carefully within the scope of the e-SIDES project.

The following initiatives and research projects are considered particularly relevant from the economic perspective. Some of them have already been introduced in the section on the societal perspective:

⁵⁵ <https://www.oecd.org/>

⁵⁶ <https://www.idc.com/>

⁵⁷ <http://www3.weforum.org/>

⁵⁸ <http://www.mckinsey.com/mgi/overview>

⁵⁹ OECD, Data-driven Innovation for Growth and Well-being, Interim Synthesis Report, 2014.

⁶⁰ IDC and Open Evidence. European Data Market, SMART 2013/0063, D8 — Second Interim Report, 2016.

- BYTE: The Big Data Roadmap and Cross-disciplinary Community for Addressing Societal Externalities⁶¹ (2014-2017) - The BYTE project assisted European science and industry in capturing the positive and diminishing the negative externalities associated with big data. Work conducted within the scope of the project can be useful for e-SIDES particularly with respect to the discovery of impacts and stakeholders relevant in the context of big data.
- EuDEco: Modelling the European Data Economy⁶² (2015-2018) – The EuDEco project assists European science and industry in understanding and exploiting the potential of data reuse in the context of big and open data. EuDEco deals with legal, socio-economic and technological success factors and visualises the relationships between elements of the data economy. e-SIDES may benefit with respect to impacts, stakeholders and solutions.
- European Data Market Study⁶³: The European Data Market study aims to define, assess and measure the European data economy. The project is focused on developing a data ecosystem of stakeholders driving the growth of this innovative market in Europe.

Understanding the value that individuals assign to ethically sound and socially acceptable big data technologies and applications is of great importance in the context of e-SIDES. While the e-SIDES consortium is not aware of any existing research that would have focussed on the whole spectrum of the subject, there is quite some work that has focused more specifically on the value assigned to the protection of personal data. e-SIDES thus builds on this work when assessing the economic viability of privacy-preserving big data technologies and applications.

Acquisti et al., for instance, discussed at length what privacy (or rather, access to personal data) is worth.⁶⁴ They found that individuals assign markedly different values to the privacy of their data depending on how much money they would accept to disclose otherwise private information, or how much they would pay to protect otherwise public information.

Understanding the value assigned by individuals is important to developers of future big data technologies because, by estimating how much individuals value effort put into addressing ethical and social challenges, managers can seek to predict which accommodating interventions may become sources of competitive advantage and which intrusive interventions may trigger adverse reactions. e-SIDES aims to help addressing the identified issues without losing sight of the economic viability of the technologies and applications.

8.3. Terminology

Cost-benefit analysis

Is a process by which business decisions are analyzed⁶⁵. The benefits of a given situation or business-related action are summed, and then the costs associated with taking that action are subtracted.

⁶¹ <http://new.byte-project.eu/>

⁶² <http://data-reuse.eu/>

⁶³ <http://www.datalandscape.eu/>

⁶⁴ Acquisti, A., John, L., Loewenstein, G., *What is privacy worth?*, The Journal of Legal Studies, Vol. 42, No. 2, 2013, <http://www.heinz.cmu.edu/~acquisti/papers/acquisti-privacy-worth.pdf>.

⁶⁵ <http://www.investopedia.com/terms/c/cost-benefitanalysis.asp#ixzz4e81gwAt2>

Sometimes a monetary value is put on intangible items, such as the benefits and costs associated with living in a certain town. It is not uncommon to also factor opportunity cost into such equations.

Economic viability

Means that a market operation is sustainable regarding current and projected revenues⁶⁶. The revenues will be greater than or equal to all current and planned expenditures. In simple terms, any project or activity that can financially support itself is economically viable.

Network effects

Effect that a single person using a good or a service has on the total perceived value of that product or service for others. The more who use the product or service, the higher its value becomes to the group.⁶⁷

Market concentration

Measures the extent of domination of sales by one or more firms in a particular market. The market concentration ratio is measured by the concentration ratio.⁶⁸

Multi-sided markets

Markets that link two or more distinct but interdependent groups of customers.⁶⁹

⁶⁶ <https://www.reference.com/world-view/economic-viability-6d24368d0bb855a3#>

⁶⁷ <http://www.businessdictionary.com/definition/network-effects.html>

⁶⁸ <http://economictimes.indiatimes.com/definition/market-concentration>

⁶⁹ Evans, D.S., *Managing the Maze of Multisided Markets*, Strategy & Business, 2003/32.

9. Conclusions

The general assumption of e-SIDES is that big data technologies offer very promising opportunities, but that they may have several negative side effects when not designed or used properly. Hence, e-SIDES will not only look at the technologies, but also at how they are used. Maximising the benefits of big data technologies while minimising negative side effects calls for privacy-preserving big data technologies. The main focus of this project is to develop a coherent assessment and validation framework that will help to develop privacy-preserving technologies that are inspired by the principles of RRI and respond to the main societal and ethical challenges raised by big data.

To achieve this goal, the study team has developed a methodological approach based on the analysis and assessment of the main challenges in a process of interaction and networking with the main stakeholders' communities and the research projects, leading to the development of shared recommendations and priorities. Thus, e-SIDES is based on an inclusive approach where ethical, legal, societal and economic challenges related to big data technologies and their application are discussed and encountered jointly with the stakeholders affected by or involved in the development of the technologies, aiming to promote the development of big data technologies that are both sustainable and responsible.

10. Glossary

Anonymizer

Privacy Enhancing Technology (PET) targeted at anonymizing personal data.

Artificial intelligence (AI)

Science focusing on mimicking human intelligence.

Big data

Datasets that are so large or complex that traditional data processing technologies are inadequate to deal with them. Big data is often characterized by 3 Vs: Volume, Velocity and Variety, sometimes complemented with additional Vs, such as Veracity and Value.

Big data technologies

Information technologies for storing, processing, analyzing, disseminating and managing big data.

Big data analytics

Big data technologies focusing on analyzing big data.

Business intelligence

The use of business information for business strategies and decision-making.

Cryptography

Encryption of information to prevent unauthorized people from accessing this information.

Cloud computing

Internet-based computing in which data storage and processing resources connected to the internet are shared.

Databases

Digital archives with large amounts of data that can be searched in automated ways.

Data controller

Pursuant to Article 3(8), it means the competent authority which, alone or jointly with others, determines the purposes and means of the processing of personal data.

Data coupling/Data transfers

Coupling personal data files of individuals from different databases (e.g., financial data, health data, passenger data, etc.) in order to obtain a more complete picture of that person.

Data fusion

Process of transforming unstructured datasets and/or data from heterogeneous sources into structured, homogeneous formats.

Data mining

Automatic analysis of data using mathematical algorithms, in order to find new patterns and relations in data.

Data recycling

Using data several times for the same purpose in the same context

Data repurposing

Using data for different purposes than for which they were initially collected, but still in the same context as the original purpose.

Data recontextualization

Using data in another context than in which they were initially collected.

Data warehouses

System for storing, reporting and analyzing data.

Deterministic record linkage

Entity resolution in which a minimum number of identifying data has to match.

Discrimination-aware data mining

Data mining technologies with built-in requirements (based on privacy by design) to avoid discovery of discriminating patterns.

Entity resolution

Determining which persons are the same in different datasets.

Image processing

Automated processing and analysis of large amounts of camera footage, for instance, with the use of facial recognition or pattern recognition.

Information technology

Systems and applications for collecting, storing, processing, disseminating and managing data.

Internet of Things

Internet of connected smart devices, including vehicles, buildings and electronic devices.

Filters

Filters can be used to scan data input and decide data output. Filters can be based on norms, profiles, models, blacklists, etc.

Fuzzy matching

Entity resolution using both identifying and non-identifying data.

Machine learning

Area in computer science that focuses on enabling computers to learn without being explicitly programmed.

Network analysis

Analysis to clarify relations between people, including 'who is who?' (entity resolution) and 'who knows whom?'

Pattern recognition

Technologies for automated data analysis. See also data mining.

Privacy by design

Approach in which privacy requirements are taken into account throughout the entire engineering process.

Privacy Enhancing Technologies (PETs)

Technologies that protect and improve privacy, particularly technologies that comply with privacy and data protection laws.

Privacy-preserving data mining

Data mining technologies with built-in requirements (based on privacy by design) to preserve privacy.

Probabilistic record linkage

See: Fuzzy matching.

Profiling

Ascribing characteristics to individuals or groups of people on the basis of available data and results of data analyses.

Record linkage

See: Entity resolution.

11. References

- Aquisti, A., John, L., Loewenstein, G. (2013), What is privacy worth?, *The Journal of Legal Studies*, Vol. 42, No. 2
- Bakhshi, H., A. Bravo-Biosca, and J. Mateos-Garcia (2014), Inside the datavores: Estimating the effect of data and online analytics on firm performance, Nesta, March, www.nesta.org.uk/sites/default/files/
- Custers, B.H.M. (2013), Data Dilemmas in the Information Society, in: Custers B.H.M., Calders T., Schermer B., Zarsky T. (red.) *Discrimination and Privacy in the Information Society*. nr. 3. Heidelberg: Springer
- Custers, B.H.M., Calders, T., Schermer, B., and Zarsky, T. (eds.) (2013), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, Heidelberg: Springer
- EU Network of Independent Experts on Fundamental Rights (2006) *Commentary of the Charter of Fundamental Rights of the European Union*, June 2006, available at: ec.europa.eu/justice/fundamental-rights/files/networkcommentaryfinal_en.pdf
- Evans, D.S. (2003), *Managing the Maze of Multisided Markets*, *Strategy & Business*, 32
- Gandomi, A., & Haider, M. (2015), Beyond the hype: Big data concepts, methods and analytics, *International Journal of Information Management*, 35
- Gavison, R. (1980), Privacy and the limits of law , *Yale Law Journal* 89 : 421–71
- Henkin, L. (1989), The Universality of the Concept of Human Rights, *The ANNALS of the American Academy of Political and Social Science*, Vol 506, Issue 1
- Hosseini, M., Renn, O., Vanclay, F. et al. (2013), A framework for combining social impact assessment and risk assessment, *Environmental Impact Assessment Review*, Vol. 43, No. 1
- IDC and Open Evidence. *European Data Market, SMART 2013/0063, D8 — Second Interim Report*, 2016
- Laney, D. (2001), *3D Data Management: Controlling Data Volume, Velocity and Variety*, Gartner. Stamford CT: META Group Inc.
- Lynskey, O. (2015), *The Foundations of EU Data Protection Law*, OUP
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburg, C. and Hung Byers, A. (2011), *Big Data: The next frontier for innovation, competition and productivity*, Seoul: McKinsey Global Institute
- Mayer-Schönberger V. & Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think*, New York: Houghton, Mifflin, Harcourt Publishing Company
- Metcalf, J., Keller, E. F., and boyd, d. (2017), *Perspectives on Big Data, Ethics, and Society*, Council for Big Data, Ethics, and Society, accessed March 20, 2017, available at: <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/>

Mittelstadt, B., Floridi, L. (2015), The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts, *Science and Engineering Ethics*, published online 23 May 2015

Nissenbaum, H. (2010) *Privacy in Context*, SUP

OECD, *Data Driven Innovation for Growth and Well-Being*, October 2015

Pollicino, O., Kowalik-Bańczyk, K. (2016), Migration of European judicial ideas concerning jurisdiction over Google on withdrawal of information, *German Law Journal*, Vol. 17 No. 03

Post, R. (1989), The social foundations of privacy: Community and self in the common law. *Tort* , 77 *California Law Review* : 957–1010

Tene, O. and Polonetsky (2012), *Privacy in the Age of Big Data; A Time for Big Decisions*, 64 *Stanford Law Review Online*, Feb 2

Vanclay, F. (2003), *International Principles for Social Impact Assessment*. *Impact Assessment & Project Appraisal* 21(1), 5-11

Vanclay, F., Esteves, A.M., Aucamp, I. And Franks, D. M. (2015), *Social Impact Assessment: Guidance for assessing and managing the social impact of projects*, International Association for Impact Assessment, Fargo ND

Warren, S., Brandeis, L. (1890), *The Right to Privacy*, 4 *Harv. L. Rev.* 193

Westin, A. (1967), *Privacy and Freedom* , New York : Atheneum

Woodcock, C. (2014), *Big data is a lot like teenage sex*, *Tech City News*, 7 August

Wright, D., Friedewald, M. (2013), *Integrating privacy and ethical impact assessment*, *Science and public policy*, Vol. 40, No. 6

Wright, D., Kush Wadhwa, K., Lagazio, M. et al. (2014), *Integrating privacy impact assessment in risk management*, *International Data Privacy Law*, Vol. 4, No. 2