D5.3
Community Position Paper (CPP)

Towards a more accountable big data governance and responsible innovation of privacy-preserving technologies

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

e-SIDES – Ethical and Societal Implications of Data Sciences

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 analyzing, 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 program 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.

D5.3 Community Position Paper

Work package	WP 5 – Validation framework
Lead author	Karolina La Fors (Leiden University)
Contributing authors	Daniel Bachlechner (Fraunhofer ISI) Alan M. Sears (Leiden University) Bart Custers (Leiden University) Michael Friedewald (Fraunhofer ISI)
Internal review	Gabriella Cattaneo (IDC) Richard Stevens (IDC)
Due Date	M33 (September 2019)
Date	20 February 2020
Version	2.0
Type	Report
Dissemination level	Public

This document is Deliverable 5.3 of Work Package 5 of the e-SIDES project on Ethical and Societal Implications of Data Science. e-SIDES is an EU funded Coordination and Support Action (CSA) that complements Research and Innovation Actions (RIAs) on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies and providing a broad basis and wider context to validate privacy-preserving technologies. All interested stakeholders are invited to look for further information about the e-SIDES results and initiatives at www.e-sides.eu.

Grant Agreement number: 731873

Executive Summary

The primary aim of this document is to map challenges and opportunities of big data stakeholders in respect to designing, implementing and using privacy-preserving technologies within big data contexts. Although privacy-preserving technologies[1] have a huge potential in adding value to the big data economy that increasingly capitalises upon how data can be used, reused, and exploited for multiple purposes by fostering citizen-centric responsible innovation. According to estimations of IDC, the big data growth will reach 175 trillion zettabytes[2] by 2025, unleashing further potentials for data science and analytics. Yet, this community position paper shows that challenges emerge when accommodating different economic, technological societal and policy-related interests of big data stakeholders. The amount of data breaches, ransomware, and cyber security attacks are growing and taking up forms, for instance, that are increasingly noticeable, and although being deployed locally can have global effects.[3] According to OECD[4] data, by 2019 cybercrime is expected to cost businesses over 2 trillion USD and citizens remain vulnerable at protecting their own data.[5] Given the conflicts between a wide economic impetus for generating and exploiting more and more data and the increasing vulnerabilities in cyberspace, this community position paper offers an outlook upon the core challenges big data stakeholders face. The paper outlines the potential but also the core challenges and limitations of responsibly innovating privacy-preserving technologies that could function as data security measures for more accountable big data solutions.[6] Moreover, this community position paper also offers an outlook upon potential opportunities that can also be regarded as recommendations for improving big data governance structures in a manner that big data innovation, including the innovation of privacy-preserving technologies, becomes more responsible and accountable. The

[1] In this community position paper, we understand privacy-preserving technologies to be technologies by which the integration of legal privacy safeguards is operationalized in big data solutions. Privacy-preserving technologies serve the purpose of preventing and limiting the unintended violation of the privacy of individuals. The right to privacy under the right to respect for private life (Art. 7) and the right to data protection (Art. 8) are separate fundamental rights preserved by the European Charter of Fundamental Rights. The European Human Rights Convention (under Article 8) also preserves the right to respect for private and family life, correspondence and home.

[2] IDC: Expect 175 zettabytes of data worldwide by 2025 – Retrieved on 23rd of May 2019 from <https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

[3] Enhancing the role of cyber insurance management - The cyber insurance market: Responding to a risk with few boundaries (OECD, 2018) - <https://www.oecd.org/pensions/The-cyber-insurance-market-responding-to-a-risk-with-few-boundaries.pdf>

[4] OECD is the Organization for Economic Cooperation and Development <https://www.oecd.org/about/>

[5] Enhancing the role of cyber insurance management - The cyber insurance market: Responding to a risk with few boundaries (OECD, 2018) - <https://www.oecd.org/pensions/The-cyber-insurance-market-responding-to-a-risk-with-few-boundaries.pdf>

[6] Under big data solutions in this community position paper we understand tools, technologies and programmes and processes as well.

Grant Agreement number: 731873

last includes a combination of ethical, legal, societal, and economic opportunities resulting from the involvement of multiple stakeholders.

By bringing together the viewpoints of different members of the big data community who are often isolated from each other, by this community position paper we attempted to build bridges between their views and relied upon them in order to reflect upon the challenges and opportunities of shaping more accountable big data governance and privacy-preserving technology innovation.

Four main challenges discussed in the community position paper are:

- *1) differing attitudes towards privacy and reusability of data and privacy-preserving technologies;*
- *2) transparency, raising awareness versus cognitive overload of users;*
- *3) discrepancies between legal compliance and ethics;*
- *4) difficulties of conducting data protection impact assessments*.

Four main opportunities for stakeholders to overcome challenges are:

- *1) modes of raising awareness and increasing transparency;*
- *2) tools of accountability;*
- *3) reference points of accountability;*
- *4) bodies and mechanisms of oversight.*

Grant Agreement number: 731873

Contents

Abbreviations

EU	European Union
BDVA	Big Data Value Association
EDPB	European Data Protection Board
PIA	Privacy impact assessment

Grant Agreement number: 731873

1. Introduction

This section outlines the background, the methodology and the structure of this document.

1.1. Background

This report is Deliverable 5.3 of the e-SIDES project. In this project, the ethical, legal, societal and economic implications of big data applications are examined in order to complement the research on privacy-preserving big data technologies (mainly carried out by ICT-18-2016 projects) and data-driven innovation (carried out, for instance, by ICT-14-2016-2017 and ICT-15-2016-2017 projects).

This deliverable provides the e-SIDES community position paper on accountability issues related to the broader governance of big data from the position of privacy-preserving technologies. This deliverable leverages upon the developers and operators of big data solutions, developers of privacy-preserving technologies, policy makers and the civil society are taken into account.

1.2. Methodology

This community position paper had been initiated by the e-SIDES consortium[7] and aimed at including the voices of a broad variety of stakeholders within the big data community in order to incentivise mutual dialogue about the main challenges faced in respect to big data innovation.

Although the main focus of the e-SIDES project is on Europe, given that influential big tech companies are located and operating from the US, we also include challenges stemming from EU-US regional differences. From the legal perspective in our methodology we focus on data protection law and to a limited extent on competition law. In our methods, we relied upon the views of experts who attended our e-SIDES workshop on the 2nd of April 2019,[8] as well as those who joined the e-SIDES community online and left a contribution on a relevant topic through our website. Among the contributors were representatives of industry, academia, policy makers, standards bodies, trade associations, data protection authorities, investors and civil society. Moreover, we also relied upon our own desk research in order to complement the list of challenges and opportunities and initiate further dialogue upon them by the community.

[7] e-SIDES project: <https://e-sides.eu/e-sides-project>

[8] e-SIDES workshop: Towards Value-Centric Big Data: Connect People, Processes and Technology, held on the 2nd of April 2019, in Brussels - <https://e-sides.eu/resources/towards-value-centric-big-data-e-sides-workshop-report >

1.3. Structure

The community position paper has the following structure:

- In section 1 the background, the methodology and the structure of the deliverable is outlined.
- In section 2, key groups of big data stakeholders are presented.
- In Section 3, challenges that are faced by or that affect big data stakeholders are presented.
- In section 4, opportunities for action that may be useful to overcome the challenges are listed and explained.

2. Big data stakeholders

Big data stakeholders bring together different and, at times, conflicting ethical, legal, societal and economic viewpoints. This provides ideal conditions for nurturing a constructive dialogue. This community position paper reflects the views of the following groups of big data stakeholders:

1. **Industry:** includes representatives of organisations developing big data solutions, the security industry working on privacy-preserving technologies, etc.; members of this group play a key role in related H2020 projects; at the company level, data protection officers (DPOs) and ethics officers play a key role alongside management;

2. **Academia:** includes researchers focusing on data technologies, ethical and societal aspects of data use, privacy-preserving technologies, etc.; includes educational institutions of all kinds; members of this group play a key role in related H2020 projects;

3. **Policy makers:** includes policy makers at the EU level (e.g., members of the European Commission) and policy makers at the national level; members of this group may focus on economic, innovation, research or education policy; includes national research funding agencies;

4. **Standards bodies:** organisations that develop, coordinate, promulgate, revise, amend, reissue, interpret or otherwise produce technical standards;

5. **Trade associations:** founded and funded by organisations operating in a specific industry (e.g., the Big Data Value Association - BDVA); they focus on collaboration between organisations, advertising, education, lobbying and publishing;

Grant Agreement number: 731873

6. **Data protection authorities:** independent public authorities that supervise, through investigative and corrective powers, the application of data protection law (e.g., European Data Protection Board (EDPB) [formerly the Art. 29 WP], national and provincial authorities);

7. **Investors:** individuals or entities that provide support which is not limited to financial investments; support may include connections, mentorship, training and public recognition;

8. **Civil society:** includes individuals who are data subjects in cases whose personal data is used but also individuals whose lives are significantly affected by certain uses of data; members of this group are usually represented by civil society organisations;

9. **Media/journalists:** includes members of the journalistic community writing about big data and using big data as commercial parties to profile readers.

3. Challenges

Big data stakeholders are facing or affected by a broad range of challenges related to ethical, legal, societal and economic aspects, which stem from the complexities data aggregation, exploitation and security pose. In this section, we dive into four main challenges regarding not only privacy and privacy-preserving technologies, but also governance mechanisms currently shaping big data structures within which privacy-preserving technologies are also situated. The four challenges that are discussed in this section are: *1) differing attitudes towards privacy and reusability of data and privacy-preserving technologies; 2) transparency, raising awareness versus cognitive overload of users; 3) discrepancies between legal compliance and ethics; 4) difficulties of conducting data protection impact assessments*.

3.1. Differing attitudes towards privacy and reusability of data and privacy-preserving technologies

1) Privacy: The stance of industry towards privacy shall change, because they either care a lot or do not care at all as to what happens beyond privacy when they balance their security interests. For those who do not care, the protection of individual rights has not yet become established as a priority. Privacy preservation is not yet the “normal” setting for operations, and privacy needs to be more embedded in organisational thinking. What would be needed is a synergetic, embedded business approach towards privacy; shields and safeguards need to be laid out so that organisations know what (and how) they should do. The question is if organisation really accept the responsibility for how they organise privacy within their

Grant Agreement number: 731873

domain (for instance, sandbox regulation is an interesting approach but so far has not brought fruitful results).

2) Usable and reusable privacy-preserving technologies: A significant challenge for developers is to create tools with privacy-preserving technologies that can be used and reused. They face challenges with respect to finding a sweet-spot of being able to develop big data solutions that can be customised but are also generic enough so that their solutions can also be reused. Reusability is also impeded by differing requirements that are related to data, staff and technologies. The requirements that data must meet differ from one context to another. Assessing data quality as well as the fit of a big data solution for a context is difficult. Yet, limiting potential biases is crucial. Therefore, it is important to know what kind of data is used and reused for which different purposes and on which legal grounds. Reusing or repurposing data might have negative consequences, because the data may not (entirely) fit to the new context. Furthermore, the adoption of technologies requires special expertise and technologies cannot easily be taken from one context to another, which hampers reusability: software is usually not context-aware, because the requirements and the technical possibilities change continuously. There remains a big gap between the supply side and the demand side because the maturing of privacy-preserving technologies requires similar commitment from industry, academia and policy around developing and implementing such technologies in different contexts. Starting more funding programmes for start-ups to explore the market of privacy-preserving technologies and bring further the (design and use of) protypes developed under the EU 7th Framework Programme would also benefit the maturation of privacy-preserving technologies.

3.2. Transparency, raising awareness versus cognitive overload of users

Big data stakeholders agreed that keeping processes transparent is crucial and raising awareness among users remains essential, but highly challenging. Experts also pointed out that consumers regularly do not understand or know the models behind data analytics, and in daily life they want less and less data management tasks because they already experience cognitive overload. This would require better data protection management structures in general and around privacy-preserving technologies. The latter may grow into an even bigger problem in the future as the data overload and the rapidly increasing amount of analytic potential makes it more difficult to make users aware of the implications of profiling. Nevertheless, users do not only have to be informed but also empowered to exercise their rights. Lack of transparency undermines trust in big data solutions as well as in developers and other stakeholders using big data solutions. Social cooling and chilling effects are among the consequences of this.

Grant Agreement number: 731873

3.3. Discrepancies between legal compliance and ethics

1) Both law and ethics lag behind technological developments and EU policy makers and academics argued that ethics needs to better accommodate and advocate for new privacy-preserving technologies, for instance, through demonstrating best practices. Such accommodation and advocacy is desirable, as going beyond legal compliance would include upholding the broader ethical definition of fairness and responsible innovation as main goals. Others, NGO's, Data Protection Authorities argued that ethics, best practices and self-regulation of companies have not worked in protecting data subjects against loss or misuse of their data and called for more regulation.
2) Ethical and legal compliance always come at a price in terms of regional differences as well. For instance, the Los Angeles Times disabled access to European residents because they did not comply with the GDPR. Beyond economic goals, this move could be regarded as being ignorant towards ethical values that facilitate the sharing of knowledge and the right to freedom of expression. Furthermore, it adds to discrepancies between legal compliance and ethics, as such discrepancies cannot be addressed by technologies alone; technologies have to be embedded in organisational processes and structures in an ethically informed and legally compliant manner.
3) The principle-based approach, based on the history of the GDPR and privacy legislation and the spirit of the law, can also provide direction in addressing discrepancies. Between the GDPR and all of the other human rights instruments, there is enough legislation but enforcement lags behind. However, DPA's and other watchdogs are slowly catching up on this, thus the need for investment in data protection matters steadily increases for the industry complex.

   For instance, the value of human dignity is key—this aspect is taken into consideration with regards to the use of genetic data and it is conceptualised by the law. The question was also raised: do we want to (or can we!) incorporate ethical values such as social cohesion or sustainable development in legislation on big data innovation in order to address discrepancies and render innovation and big data governance more responsible?

3.4. Difficulties in conducting evaluations and data protection impact assessments

1) Difficulties in assessing the impact of data and complying with the GDPR in general are a challenge, because 1) people do not even know what questions to ask; 2) this leads to a low demand for privacy-friendly and ethical solutions; 3) a low demand leads to limited efforts to find solutions. The assessment of risks was genuinely regarded as being necessary (context-dependent) and proportionate. The principle of proportionality during risk assessments should lead to scoping and need to be applied against certain benchmarks. Such benchmarks shall include a list of ethical values that are implicated and problems that need to be addressed by data processing.

Grant Agreement number: 731873

2) Data protection impact assessments currently do not involve the identification of the applicable parts of the ePrivacy Directive and other relevant legislation. A strong body of oversight is missing in order to evaluate compliance. But the question remains as to which bodies should evaluate compliance with human rights?

4. Opportunities for stakeholders to overcome challenges

There are numerous opportunities to overcome the challenges that big data stakeholders are facing. In this section, opportunities are outlined by putting particular emphasis on how the challenges raised earlier can be addressed. This section describes four clusters of opportunities each of which has to do with improving accountability in big data structures and facilitating responsible innovation: 1) *modes of raising awareness and increasing transparency; 2) tools of accountability; 3) reference points of accountability; and 4) bodies and mechanisms of oversight.* Each of these four clusters of opportunities facilitates different methods for increasing accountability; however, combining all four is vital in order to achieve more accountable big data governance and responsible innovation of big data and privacy-preserving technologies. Emphasis will be put upon what can be initiated and utilized by big data stakeholders.

4.1. Modes of raising awareness and increasing transparency

Transparency helps creating trust among stakeholders as it must always be clear how certain decisions were made. The following opportunities could contribute to raising awareness and transparency among users:

1) Collective platforms could facilitate the dialogue between a diversity of big data stakeholders but most importantly would involve the representation of users and could be initiated by users themselves. Offering a voice for these diverse stakeholders on different platforms could assist in rendering processes more accountable.

2) Economic incentives could support awareness raising, if they would also economically incentivise citizens to become involved in creating economic value of data. In return citizens could use their strengthened position as a check upon other stakeholders exploiting their data. However, different types of economic models for privacy-preservation could be realised. For instance, new business models could be considered that would render a data subject into a certain form of 'shareholder' of the profit that stems from their data. Such an arrangement would reward consumers economically as well, as current business models rely upon data that is owned by others. The majority of currently existing business models rely upon personal data as a currency in exchange for free services and these business models are doomed with respect to privacy violations. Therefore, economically incentivising the data subject can offer real benefits.

Grant Agreement number: 731873

3) Cutting out the middle-man (such as third-party trackers) could also increase the transparency of data processing for users. Users may be provided with access to their data in such a way that at the same time protects them from cognitive overload. But in general, the traceability of data for users is essential (cf. data provenance) as inferred data may be erroneous.

4) Education as a holistic approach towards multiple stakeholders (users, developers, academics) is also considered a beneficial method to raise awareness. Citizens need to take ownership, yet they should also be protected from cognitive overload; possibilities should be granted for a certain form of participatory regulation where citizens' participation is granted – e.g.: through trade unions and how they interact with regulation that affects their trade.

5) Making privacy-friendliness part of marketing activities (e.g., "We do the right thing.", "Do no evil.") can also be a helpful awareness-raising tool as it may allow for competitive differentiation and increasing the demand for privacy-friendly solutions.

All these awareness-raising modes that are initiated by different stakeholders depending on their position in the big data value chain shall empower data subjects not only about their rights but also about broader accountability mechanisms and about their position within big data governance structures. Raising awareness shall also include awareness about biases: a system itself may not be discriminatory but a system 'learns' to discriminate from members of society which feed information into it that can be discriminatory (systems can be regarded as mirroring and even amplifying a large part of societal perceptions). Specifying who is accountable for what; particularly when decisions are taken with the help of a big data solution is essential.

4.2. Tools of accountability: Licensing, standards, incentives, certifications and others

1) A licensing and ethical framework for developers that is validated by a diverse range of experts, potentially by an oversight body, may be useful. Such licensing may be something that can bring different attitudes to privacy by different stakeholders and can facilitate the reusability of data.

2) Sector-specific standards may also be essential, and these would also take into account ethical values and assess fairness, starting with the public sector and moving towards the private sector. In general, standards make it easier for stakeholders to assess what had been done by another stakeholder. The development and application of independent sector-specific ethical frameworks may also be beneficial to underpin standards.

Grant Agreement number: 731873

3) Certifications shall be developed with the involvement of industry and a broad representation of society, including independent experts but also citizens' assemblies.[9] How a certain certification comes about, from which stakeholders and what to certify: processes, outcomes, governance models or what other parts of the big data value chain shall be made transparent. Ethics or privacy seals as forms of certification are also perceived as beneficial, yet ethics seals are complex to arrange because they need to gain general trustworthiness among the broad public including end users.

4) Technology-embedded 'data borders' and 'passports' for data subjects may also be a form of standard in order to maintain control in the hands of data subjects. But also these tools would limit access by rendering context-dependent limitations. But how such borders and passports would achieve that and how to organise democratic oversight above these two tools of user control remains unclear.

5) Big data as metadata (aggregated and anonymised data) may also be a way to ensure user control and that software avoids the misuse of data proactively (e.g., Photoshop detects banknotes).

6) Risk assessments during different phases of innovation and across different contexts: Research has demonstrated that risk assessments shall accompany not only the design-phase or only the application-phase of technologies but assessments of potential value violations shall accompany the circular life cycle of big data where design, application and re-design are phases that regularly overlap.[10] Therefore, and because data as a basis for analysis is agile, models change; big data solutions need to be flexible and privacy-preserving technologies need to be adaptable; evaluations shall not only focus on separate phases of design, application and redesign are needed for improvement. Furthermore, risk assessment shall also encompass how data is used and reused across contexts. Health data, for instance, needs special protection; ethical and GDPR-compliant technological solutions would be useful for sharing data across contexts. In designing and conducting privacy or data protection impact assessments, the best practices of public institutions should also be followed. When conducting Privacy Impact Assessments (PIAs) in Canada, PIAs must be attached to all data exchanged in Canada. If the assessment leads to negative results, the

[9] A powerful spyware app now targets iPhone owners - https://techcrunch.com/2019/04/08/iphone-spyware-certificate/

[10] La Fors, K., Custers, B. & Keymolen, E. (2019) Reassessing values for emerging big data technologies: integrating design-based and application-based approaches, Ethics Inf Technol (2019). https://doi.org/10.1007/s10676-019-09503-4

systems will not be funded. This mechanism is in place since the mid-90s in Canada and the EU could requests assessments in a similar fashion with respect to public procurement.

4.3. Reference points of accountability: Legislative opportunities, court cases and ethical codes of conduct

1) Arguments for capitalising upon or amending current legislation and developing additional ones all co-exist. New piece of legislation on the Internet of Things in general is needed, and the Directive on the legal protection of computer programs[11] must be broadened in terms of its scope. The General Product Safety Directive[12] should also be amended as it currently does not recognise software as a product. Therefore, data analytics currently falls in a legal vacuum as the context of self-driving cars, for instance, also demonstrates. There are many rules but they need to be updated for the big data and analytic requirements of the 21st century.

2) Court cases are also necessary to interpret the GDPR, because the text of the legislation is difficult and clarification would be very helpful for companies regarding its interpretation and also implementing it into product design. The European Court of Justice ruled for instance in the Planet49 case[13] that active consent is required from users. The judgement came as a result of a lawsuit filed by the German Federation of Consumer Organisations against the use of a pre-ticked checkbox which had consented to cookies on behalf of the user by default. The judgement clarified that opt-out options do not constitute meaningful consent under the GDPR and the ePrivacy Directive. Another important judgment in the Fashion ID GmbH & Co. KG[14] clarified the concept of joint controllership. CJEU decision specified that such website operators implementing the Facebook 'Like' button, like Fashion ID have to inform data subjects of the collection of personal data and the transfer of such data to Facebook. Furthermore, the judgment also

[11] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Text with EEA relevance) - <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009L0024>

[12] Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety (Text with EEA relevance) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32001L0095>

[13] Request for a preliminary ruling from the Bundesgerichtshof (Germany) lodged on 30 November 2017 — Planet49 GmbH v Bundesverband der Verbraucherzentralen und Verbraucherverbände — Verbraucherzentrale Bundesverband e. V. (Case C-673/17) - <http://curia.europa.eu/juris/document/document.jsf?text=&docid=200625&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=2319411>

[14] CJEU C-40/17 (2018) Opinion of AG Bobek Case C-40/17 - <http://curia.europa.eu/juris/document/document.jsf?text=&docid=209357&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=6652452>

includes the obligation towards operators that where necessary under Directive 2002/58/EC they must ask for consent before data transfer occurs. Another important element of this judgment is that such website operators which implement the like button from Facebook are only responsible and liable for the data transfer until it arrives at Facebook, they are not liable for Facebook's data processing to other parties.

3) Beyond such cases judgments of the European Court of Human Rights that reflect upon the effects of AI, such as discrimination or other (un)intended biases, would also be helpful.[15] They can in general also be beneficial to offer companies help in getting a grip on the GDPR and to understand rights. For this, current and future court cases would be highly beneficial to aid in interpretation. But they can also be helpful in assisting citizens to learn what the GDPR offers in their daily lives as a private citizen in relation to public authorities and as users or customers in relation to private parties.

4) Investing in more legal informatics systems seems also vital. Translating the GDPR prescriptions into technical requirements appears difficult due to their vagueness. Even lawyers give different answers to one and the same question. There appears to be a sliding scale from 'it's fine' to 'no' as there is too much data. Help for lawyers is needed too. For instance, legal knowledge graphs could offer assistance to represent all possible cases in a machine-readable format, including on cases of privacy violations. This would allow comparing a much broader range of cases by the help of a multidisciplinary team.

5) Ethical codes of conducts should be revisited and developed with an emerging big data-related focus and sector-specific ethics codes, for instance, for programmers and data scientists.

4.4. Bodies and mechanisms of independent oversight

1) Strong, independent bodies of oversight are needed in order to evaluate compliance with ethical values and the legal framework. This comes also in line with latest developments regarding difficulties in maintaining ethics boards as accountability remains a major problem. The first New York City Automated Decision Systems Task Force,[16] for instance, struggles with getting access to the actual AI systems used by city authorities.[17] Controversies around ethics boards remain an issue, because ethics boards do not have regulatory powers and their composition can remain a politically difficult issue, as

[15] ECHR 101 (2019) Mart and Others v. Turkey (Application No. 57031/10) ECtHR

[16] https://ainowinstitute.org/nyc-ads-task-force-letter-re-public-engagement-030119.pdf

[17] New York cities algorithm task force is fracturing - <https://www.theverge.com/2019/4/15/18309437/new-york-city-accountability-task-force-law-algorithm-transparency-automation>

demonstrated by the dissolution of Google's Ethics Board after a week of existence.[18] Therefore, those bodies that can offer proper oversight need to have certain forms of democratic power that grants them regulatory effectivity by enforcement and consequently their appointment shall go through some form of parliamentary processes.[19] It was also suggested that a Ministry of Digitisation should be established as an overseeing body, but also as an active contributor of policy on responsible innovation of big data solutions, including privacy-preserving technologies.

2) Oversight needs to be robust; an ethics board or an ethics officer is needed that has the ability to judge and the power to act based on the judgement; a DPO usually just focuses on the data but there is more to be taken into account; ethics committees should be influential by initiating, composing, and maintaining them through a diverse combination of expertise and interests. Trusted entities that can obtain clearance from local DP offices and professional bodies for programmers can exemplify independent bodies of oversight.
3) A system of protection, rewards and incentives for EU companies that respect ethical standards could be offered by national government and supranational governing bodies. Ethicists also suggested countries could be made responsible for monitoring data ethics.
4) Oversight bodies need to have answers on how to inspect big data related products, services and organisations in a manner that is accountable, particularly to the public; the continuity of monitoring is crucial. Such methods for accountability could be supported by company audits and the publication of independent reviews of companies. As a traditional random form of auditing, the practices of investigative journalists and members of civil society remain crucial.

[18] Statt, N. - Google dissolves AI ethics board just one week after forming it (4th April 2019) - <https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-ends-controversy-kay-coles-james-heritage-foundation>

[19] https://www.technologyreview.com/s/613281/google-cancels-ateac-ai-ethics-council-what-next/

Grant Agreement number: 731873

5. Conclusions

As a conclusion, the introduced challenges and opportunities of big data stakeholders demonstrated that all relevant stakeholders need to be involved in decision-making processes and in the development accountable big data governance structures and responsible innovation of privacy-preserving technologies in different contexts. A mix of views on data use is needed (e.g., legal, technical, business). The GDPR achieved that technology experts talk to legal experts, but there are further perspectives that need to be considered; a business perspective is needed to make the risk judgement; decision makers in the companies need to be on board; involve a translator who understands the languages of the relevant perspectives (e.g., law has to be translated into technical requirements); understanding the other fields to some extent can be an alternative. Furthermore, bodies of oversight shall be established which – beyond courts – are acknowledged by a wide diversity of big data stakeholders.

Grant Agreement number: 731873