Deliverable D5.1
Assessment of Existing Technologies Under Development

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.

D5.1 Analysis of PPTs under Development

Work package	WP 5 – Validation framework
Lead authors	Karolina La Fors (Leiden University) Alan M. Sears (Leiden University)
Contributing authors	Daniel Bachlechner (Fraunhofer ISI) Michael Friedewald (Fraunhofer ISI) Jana Weitkamp (Fraunhofer ISI) Melek Akca Prill (Fraunhofer ISI) Bart Custers (Leiden University)
Internal review	Richard Stevens (IDC)
Due Date	M29 (May 2019)
Date	20 June 2019
Version	1.0
Type	Report
Dissemination level	Public

This document is Deliverable 5.1 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 visit www.e-sides.eu for further information about the e-SIDES results and initiatives.

Executive Summary

This report was built upon Deliverable 4.2 of the e-SIDES project, which developed four general requirements for the use of big data solutions: security and privacy features should be embedded in the solutions; data breaches should be avoided before they happen; solutions, processes and people should fit together; and compliance with laws and policies should be ensured. In line with this, the aim of this current report was to take a step forward and create an inventory of the impact assessment modes and tools of data-driven innovation and privacy-preserving technologies and the extent to which and how ethical, legal, social and economic challenges are considered. Therefore, this deliverable answers the research question: How are security and privacy features in big data-driven innovation projects—which we call privacy-preserving technologies—suitable to deal with the ethical, legal and societal values that come under pressure in specific big data application contexts? We also look at how the ethical, legal and societal implications of such technologies are measured.

In order to answer this research question, we first identified a set of data protection impact assessment, social impact assessment and responsible research and innovation (RRI) assessment tools, notably the following: Art. 29 WP Guidelines on DPIA,[1] Data ethics impact assessment, Data Ethics Canvas of the Open Data Institute, Software Impact Assessment; ISO/IEC 29134:2017 Information technology -- Security techniques -- Guidelines for privacy impact assessment, responsible innovation assessment tools and the Applied Ethics Toolkit developed through Deliverable 2.2 of e-SIDES.

Second, we identified three big data areas in order to assess how are values and the implications of privacy-preserving technologies and big data solutions assessed. These three areas were the following: healthcare; transportation and smart cities; and web browsing and third-party tracking. Each of these areas were chosen for a set of reasons. Healthcare was chosen due to the shrinking prices of medical examinations, the growing amount of wearable mobile devices, the increasing ageing population, and because the ever-growing demand for data and analytics has become a common phenomenon. Transportation and smart cities were chosen because the accumulation of purposes for which big data is used in smart cities has grown and recent years and looks as though it will continue on this trajectory. Web-surfing and third-party tracking was selected because big data developments in this sector have major implications for the privacy of individuals, as advertisements and other services are targeted based on personal data collected through both mobile applications and browsing the Internet.

e-SIDES is a Coordination and Support Action aiming at complementing research on and development of privacy-preserving big data technologies and data-driven innovation funded under the European Commission's Horizon 2020 funding scheme. Therefore, we mapped how the specific ethical, legal and societal implications of privacy-preserving technologies already in the design phase of technology development were assessed within a set of ICT-14, ICT-15 and ICT-16 data-driven innovation projects funded under the European Union's Horizon2020 scheme. These projects were the following: BODYPASS,[2] MyHealthMyData,[3] CLARUS,[4] BigMedilytics[5], SPECIAL,[6] AEGIS[7], and Transforming Transport[8]

[1] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
[2] http://www.bodypass.eu/
[3] http://www.myhealthmydata.eu/
[4] http://clarussecure.eu/
[5] https://www.bigmedilytics.eu/
[6] https://www.specialprivacy.eu/

and the Papaya project.[9] Furthermore, through these ICT-14, ICT-15 and ICT-16 projects we also evaluated the development of privacy-preserving technologies within these three big data contexts. For the assessment, desk research was done and project representatives were invited to fill in a questionnaire via e-mail in order to provide relevant feedback on the existing impact assessment techniques and modes within their projects. Furthermore, the extent to which requirements for the design and (to a limited extent) the use of these technologies were met by big data solutions that were currently under development were also evaluated. We also offered recommendations as to what impact assessment tools out of the seven listed impact assessment tools these projects could capitalise upon depending on the types of privacy-preserving technologies, the contexts, the accompanying processes and the goals of using these technologies.

From assessing the listed ICT-14, ICT-15 and ICT-16 projects, we concluded that the emergence of big data changes the protection of privacy as well as the relevance of other related issues, such as confidentiality and utility significantly. Moreover, the evaluation of projects also demonstrated that by compromising towards privacy, the granularity of data for business purposes becomes more difficult to exploit. Additionally, the rapid innovative developments and the variety of ways in which privacy-preserving technologies are used even in combination with each other, e.g.: in healthcare, also underline the importance of time during the assessment of impacts. From all three contexts, it became apparent that privacy-preserving technologies and data sharing methods serve the optimisation of processes that are increasingly geared toward improving predictions on the future behaviour of citizens. Hence, beyond the continuity in assessing the present impact of technologies, impact assessments shall also embrace the consideration of potential impacts in the future.

[7] https://www.aegis-bigdata.eu/
[8] https://transformingtransport.eu/
[9] https://www.papaya-project.eu/

Contents

Figures

Tables

Abbreviations

API	Application Programming Interfaces
CJEU	Court of Justice of the European Union
CNIL	The French Data Protection Authority
CSA	Coordination and Support Action
DPD	Data Protection Directive
EDPB	European Data Protection Board
ENISA	European Union Agency for Network and Information Security
EMR	Electronic Medical Records
ISO	International Standards Organization
GDPR	General Data Protection Regulation
OBD	Onboard diagnostic
OECD	Organisation for Economic Co-operation and Development
PSPS	Public Safety and Personal Security
PIA	Privacy Impact Assessment
RRI	Responsible Research and Innovation
RTLS	Real-Time-Locating Systems
SMEs	Small and Medium-sized Enterprises
SoDIS	Software Development Impact Statement
VPN	Virtual Private Networks
WP29	Article 29 Working Party

1. Introduction

1.1. Background

This report is Deliverable 5.1 of the e-SIDES project. In our previous deliverables (D4.1 and D4.2) we assessed what kinds of ethical legal, social and economic barriers impede the broader implementation of privacy-preserving technologies into big data solutions in different contexts. Deliverable 4.2 developed four general requirements for the use of big data solutions: security and privacy features should be embedded in the solutions; data breaches should be avoided before they happen; solutions, processes and people should fit together; and compliance with laws and policies should be ensured.

In this report we take a step further; we aim to assess the extent to which ethical, legal, societal and economic design requirements are relevant. To a limited extent we also perform this assessment for the use of big data solutions. For this purpose, we use a combination of data protection impact assessment, societal impact assessment and responsible research and innovation (RRI) assessment tools in order to evaluate the data-driven innovation projects funded under ICT-14 and ICT-15. More specifically, we inventory how within these projects the impact of specific privacy-preserving technologies is relevant in addressing ethical, legal, societal and economic values. This includes the assessment of several newly developed (groups of) privacy-preserving technologies implemented in different big data contexts that are relevant for data-driven innovation projects. Furthermore, the extent to which requirements for the design and (to a limited extent the) use of these technologies are or will be met by big data solutions that are currently under development. Therefore, this deliverable poses the research question: How are security and privacy features in big data-driven innovation projects—which we call privacy-preserving technologies—suitable to deal with the ethical, legal and social values that are under pressure in specific big data application contexts? The answer to this research question will also include answers on whether the embedded features add new issues and put additional values at risk. For this purpose, we include a set of ethical, legal and social impact assessment tools: Art. 29 WP Guidelines on DPIA,[10] Data ethics impact assessment, Data Eticas Canvas of the Open Data Institute, Software Impact Assessment; ISO/IEC 29134:2017 Information technology -- Security techniques -- Guidelines for privacy impact assessment, responsible innovation assessment tools and the Applied Ethics Toolkit developed within Deliverable 2.2 of e-SIDES.

By relying upon these assessment tools, it will also be evaluated whether embedded features (including privacy-preserving technologies) in data-driven innovation projects add new issues and put additional values at risk. These assessments will focus both on ethical and legal compliance regarding privacy and other human rights. The findings in this deliverable will serve the purpose of putting down the foundation upon which our upcoming Deliverable D5.2 recommendations can be formulated for improving design requirements that address ethical, legal, societal and economic issues.

[10] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

1.2. Methodology

The assessment of data-driven innovation projects in this deliverable is based on desk research and a series of interviews. We conducted interviews with 5 data-driven innovation projects of the European Union's framework programme, which included the following: BODYPASS,[11] MyHealthMyData,[12] CLARUS,[13] BigMedilytics[14], SPECIAL,[15] AEGIS[16], and Transforming Transport[17] and the Papaya[18] project.

The interviews were comprised of the following six questions:

1. Within your project what are the biggest challenges in terms of balancing your market interests in offering big data solutions versus ethical, social and legal issues that might come under pressure?
2. In case within your project privacy-preserving technologies (please indicate type) play a crucial role in addressing issues and fulfilling market-oriented innovation purposes what are the strengths and potential drawbacks in embracing privacy-preserving technologies for those purposes?
3. Have you carried out any form of assessment with respect to the impact of using your big data solutions?
4. In case privacy-preserving technologies are incorporated in your project (please indicate the type of it) have you carried out impact assessment with regards to the design and use of privacy-preserving technologies versus the innovation action goals of your project?
5. Are you aware of any best practices (e.g.: for instance, in terms of avoiding data breaches but also other issues) or failures or newly emerging risks that were related to the design, use or implementation process of privacy-preserving technologies which you have found beneficial to learn from? Why did you find them specifically useful for your own project goals?
6. Why do you think your big data solution is unique regarding innovation values versus ethical, social and legal values?

The questions were aimed at assessing whether data-driven innovation projects face already known or new ethical, legal, social or economic issues and whether they found additional values to be at risk. These questions were also used to guide desk research on ICT-14, ICT-15 projects that were not within the scope of our interviews but were relevant for the overall research question of this deliverable. The results of the interviews and of the desk research were integrated into the overall assessment of data-

[11] http://www.bodypass.eu/
[12] http://www.myhealthmydata.eu/
[13] http://clarussecure.eu/
[14] https://www.bigmedilytics.eu/
[15] https://www.specialprivacy.eu/
[16] https://www.aegis-bigdata.eu/ - The AEGIS project has cooperated with e-SIDES in several respects.
[17] https://transformingtransport.eu/
[18] https://www.papaya-project.eu/

driven innovation projects. Furthermore, the scientific literature review aimed at collecting different assessment tools from the disciplines of data protection, computer science, data ethics and also from the e-SIDES project itself. On the basis of literature review and frequency of scientific referencing we selected six different impact assessment tools that we found useful to evaluate ethical, legal, social and economic implications of the introduction of data-driven innovations. The first assessment tool is the Art. 29 Working Party Guidelines on DPIA[19] including the European Data Protection Board's (EDPB) Opinion on data protection impact assessment (DPIA) by member states; the second tool is the United Nations' Global Pulse 'Data Innovation Risk Assessment Tool';[20] the third tool is the Data Ethics Impact Assessment Tool, the fourth tool is the Data Ethics Canvas of the Open Data Institute;[21] the fifth tool is the Software Development Impact Assessment tool; the sixth tool is the ISO/IEC 29134:2017 - Information technology -- Security techniques -- Guidelines for privacy impact assessment; and the seventh tool this deliverable also relies upon is the Applied Ethics Toolkit of the e-SIDES project that was presented in Deliverable 2.2. Therefore, with this deliverable we aim to assess which values come under pressure in relation to different ICT-14 and ICT-15 data-driven innovation projects.

We have chosen three contexts in order to assess how privacy-preserving technologies are developed and integrated within big data solutions: healthcare, transportation/smart cities and web browsing & third-party tracking. We have chosen healthcare for the following reasons. First, because within the context of healthcare the diminishing costs of medical examinations in genetics, epidemiology, diagnostics, sequencing, biomedicine and the growing amount of wearable, mobile devices allowed for a massive production of sensitive data[22], which is more than 150 exabytes per year[23]. Second, given this amount of data is aimed at fostering innovation and boosting research securing such data is not only a legal obligation but it must require coordinated efforts throughout the whole big data life cycle from the data input, processing until inferences.

We have chosen for the context of transportation and smart cities, because of the increasing ubiquity and combinability of digital and autonomous technologies in (therefore also called as) smart cities allows for the exponential lengthening of the big data life cycle. Given the accumulation of purposes for which big data is used in smart cities, such as urban safety, environmental friendliness and more

[19] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
[20] https://www.unglobalpulse.org/privacy/tools
[21] https://theodi.org/article/data-ethics-canvas/
[22] Sight diagnostics launches an AI-based diagnostics device for faster blood tests - https://techcrunch.com/2018/07/12/sight-diagnostics-launches-an-ai-based-diagnostics-device-for-faster-blood-tests/
[23] My Health My Data - < http://www.myhealthmydata.eu/why-mhmd/>; Kim J. W., Jang B., Yoo H. (2018) Privacy-preserving aggregation of personal health data streams. PLoS ONE 13(11): e0207639. https://doi.org/10.1371/journal.pone.0207639, Retrieved on 10th of May 2019 from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207639

efficient transportation and city management all of which purposes are increasingly aimed to be achieved by forecasting and prediction that are based on personalised experiences of city inhabitants.[24]

As for the web browsing and third-party tracking context, it was chosen for a few reasons. The targeted advertising that is the focus of the context has major implications for the privacy of individuals, as the advertisements are targeted on the basis of personal data collected through both mobile applications and web browsing. Moreover, the field continues to rapidly grow: digital ad spending is estimated to be nearly 300 billion euros worldwide in 2019, which amounts to half the global advertising market, as well as for a number of markets in the EU.[25]

Although we could have chosen for other areas as well, these three big data contexts cover, first a wide spectrum of challenges that are from the perspective of ethics, law, societal and economic prosperity highly relevant to tackle. Second, these three areas also exemplify a broad diversity of context-dependent requirements that redefine expectations and goals even by seemingly the same privacy-preserving technologies.

As to the assessment tools, we identified 7 tools, which are internationally acknowledged and as the 8th tool we added our tools we developed for Deliverable 2.2. Given these tools serve different purposes, we do not intend to use them holistically on all ICT-14, ICT-15, ICT-16 data-driven innovation projects. But we will use these tools as a pallet to offer advice as to which project might benefit from which tool in order to improve the assessment of the implications of different (context-dependent) data-driven innovation projects.

1.3. Structure

The structure of this deliverable looks as follows. In section two we introduce the assessment tools. First we introduce the Art. 29 Working Party Guidelines on DPIA[26] including the EDPB's Opinion on data protection impact assessment (DPIA) by member states; second the United Nations' Global Pulse 'Data Innovation Risk Assessment Tool';[27] third the Data Ethics Impact Assessment Tool, fourth the Data Ethics Canvas of the Open Data Institute;[28] fifth the Software Development Impact Assessment tool; sixth the ISO/IEC 29134:2017 - Information technology -- Security techniques -- Guidelines for privacy impact assessment; and as a seventh tool the Applied Ethics Toolkit developed under Deliverable 2.2 of the E-SIDES project. Section 3 focuses on the assessment of different data-driven innovation projects (and privacy-preserving technologies) in different big data contexts. The projects under assessment include BODYPASS, MyHealthMyData, CLARUS, and BigMedylytics for the healthcare sector, and SPECIAL, AEGIS,

[24] Torre-Bastida, A. I. et. al. (2018) Big Data for transportation and mobility: recent advances, trends and challenges, IET Intelligent Transport Systems, Vol. 12, No. 8 - https://ieeexplore.ieee.org/document/8461278/citations?tabFilter=papers#citations

[25] The exact estimated amount is $333.25 billion USD. J. Enberg, , "Digital Ad Spending 2019" - https://www.emarketer.com/content/global-digital-ad-spending-2019

[26] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[27] https://www.unglobalpulse.org/privacy/tools

[28] https://theodi.org/article/data-ethics-canvas/

and Transforming Transport for transportation and smart cities. For the web browsing and third-party tracking context, privacy-preserving technologies are distinguished as to their use by stakeholder groups. Section 4 provides the conclusions.

2. Legal, ethical and social impact assessment tools

The introduction of new technologies into big data environments may impact a range of actors in different areas. Different tools have been developed in order to anticipate the impact the technologies will have. This chapter thus examines a range of impact assessment tools that cover legal, ethical and social areas, including: the Art. 29 Working parties Guidelines on DPIA (section 2.1), the United Nations' Global Pulse 'Data Innovation Risk Assessment Tool' (section 2.22.1), data ethics impact assessment (section 2.3), the Data Ethics Canvas of the Open Data Institute (section 2.4), software development impact assessment (section 2.5), ISO/IEC 29134:2017 - Information technology -- Security techniques -- Guidelines for privacy impact assessment (section 2.6), and the e-SIDES Deliverable 2.2 Applied Ethics Toolkit (section 2.7).

2.1. Art. 29 Working parties Guidelines on DPIA[29] (This will include EDPB's Opinion on data protection impact assessment (DPIA) by member states)

Article 35 of the General Data Protection Regulation (GDPR) introduced the concept of a Data Protection Impact Assessment (DPIA),[30] and Article 27 of Directive 2016/680, which applies to law enforcement, includes a similar provision.[31]

A DPIA is "a process designed to describe the processing, assess its necessity and proportionality and help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data by assessing them and determining the measures to address them".[32] These impact assessments are important for accountability; they aid controllers—who are responsible for conducting DPIAs under the GDPR—to comply with the GDPR and to demonstrate their compliance. A DPIA may concern a single data processing operation, multiple processing operations that are similar, or the data protection impact of a technology product. They are to be carried out before the processing, and it is not a one-time exercise, but rather a continual process.

If DPIA requirements are not met under the GDPR, such as by conducting a DPIA incorrectly,[33] or by failing to consult the relevant supervisory authority when necessary,[34] it can lead to administrative fines of up to 10M€, or up to 2% of total worldwide annual revenue, whichever is higher. Supervisory authorities should be consulted where the residual risks are high.

---

[29] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[30] 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 (General Data Protection Regulation).

[31] Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

[32] Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA), p. 4.

[33] Article 35(2) and (7)-(9) of the GDPR.

[34] Article 36(3)(e) of the GDPR.

Conducting a DPIA is only strictly required where a type of processing is "likely to result in a high risk to the rights and freedoms of natural persons".[35] However, even if the legal obligation is not triggered, controllers have a general obligation to implement adequate security measures to manage risks to data subjects. "In practice, this means that controllers must continuously assess the risks created by their processing activities in order to identify when a type of processing is 'likely to result in a high risk to the rights and freedoms of natural persons'."[36] Moreover, "[w]here appropriate", the controller must "seek the views of data subjects or their representatives".[37]

The minimum features of a DPIA are set out in the GDPR (Article 35(7), and recitals 84 and 90):

- "a description of the envisaged processing operations and the purposes of the processing";
- "an assessment of the necessity and proportionality of the processing";
- "an assessment of the risks to the rights and freedoms of data subjects";
- "the measures envisaged to:
    - "address the risks";
    - "demonstrate compliance with this Regulation"

Some tools have also been developed to help facilitate conducting DPIAs. One such tool is an open source Privacy Impact Assessment (PIA) software that pulls data from the GDPR, the PIA guides and the Security Guide from the CNIL (the French data protection authority), to aid in understanding the aspect being studied.[38]

2.2. The United Nations' Global Pulse 'Data Innovation Risk Assessment Tool'[39]

UN Global Pulse has developed a 'Risks, Harms and Benefits Assessment'. It is intended to be used at the beginning of a data innovation project, or when an existing project is changing. The initial test is the 'Data Innovation Risk Assessment Tool', which should aid in determining whether a more comprehensive 'Risks, Harms and Benefits Assessment' should be conducted. As the latter tool is yet to be released, this section will focus on the former.

The 'Data Innovation Risk Assessment Tool' is not based upon any particular legal framework, however it is influence by international and regional frameworks pertaining to privacy and data protection. The assessment tool provides a basic set of questions and a checklist with guiding comments, which are "designed primarily as a general example for internal self-regulation".[40] As this checklist offers only minimum guidance, users of the tool "are encouraged to expand the list depending on the project's

[35] Article 35(1) of the GDPR.
[36] Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA), p. 6.
[37] Article 35(9) of the GDPR.
[38] https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment
[39] https://www.unglobalpulse.org/privacy/tools
[40] http://unglobalpulse.org/sites/default/files/Privacy%20Assessment%20Tool%20.pdf

needs, risks, or specific context, or in response to the evolving data landscape".[41] The tool may also need to be adapted to fit the nature of the organization and applicable laws.

The 'Data Innovation Risk Assessment Tool' has users describe or identify:

- the purpose of the data use and the parties involved and their roles;
- the type of data that is implicated (e.g., personal data, personally identifiable data, or sensitive data);
- how the data will be obtained, used, and stored as well as whether data protection principles, such as legitimate aim and data minimization (among others), are met;
- whether the data innovation project is open or transparent and its level of transparency;
- whether partners with whom the data is shared or stored abide by similar levels of protection;
- whether the use of the data pose risks to individuals or groups and whether steps can be taken to mitigate those risks.

When certain questions of the above are answered in the affirmative or negative, it means a risk is present. For the final assessment, a judgment is made as to whether the risks are disproportionately high in relation to the potential benefit of the project, and whether to cancel, mitigate the severe risks and proceed, proceed as the risks are not severe, or obtain more guidance from experts (legal, privacy, security, etc.).

2.3. Data ethics impact assessment and guidelines (DataEthics)[42]

DataEthics has formulated a basic data ethics impact assessment as well as principles and guidelines to aid in integrating data ethics into data processing activities. The idea is to take "the analysis of impacts one step further than mere legal compliance. It would consider factors such as community influence, social risks, the distribution of responsibilities."[43]

The minimum questions to ask for a data ethics impact assessment concern:

- "The law: Is it interpreted from the point of view of the individual?
- Interests: Who or what does the data processing benefit? (power relations) What is the explicit rationale? (e.g. who supplies the data? And who benefits from it?)
- (AI) Training Data: Which cultural values/biases does the training data represent? Can we live with them? Are they transparent? Can the training data be manipulated in ongoing processes?
- Transparency: Can data processes be traced and explained?
- Accountability: Can the system rectify? Be audited by an external auditor?
- Human Empowerment/Agency: How can the needs and values of the communities and stakeholders affected by the data processing be met in the design?"[44]

[41] Ibid.
[42] https://dataethics.eu/en/time-data-ethics-impact-assessment/
[43] https://dataethics.eu/en/time-data-ethics-impact-assessment/
[44] Gry Hasselbalch, March 2017.

The principles and guidelines provide a more in-depth explanation and analysis of these issues.[45] The principles state that human beings should be at the centre and the primary beneficiary of the processing, individuals should have control over their data, data processing and automated decision-making should be transparent and understandable to individuals, organisations should be accountable for their processing and storage of data as well as that of partners, and processing should be equal and non-discriminatory so as to not affect vulnerable people. The principles are then followed by a questionnaire that asks a range of questions within subtopics relating to each of the aforementioned principles.

2.4. Data Ethics Canvas of the Open Data Institute[46]

The Data Ethics Canvas is intended "for anyone who collects, shares and uses data".[47] In addition to aiding legal compliance, the aim of the Canvas is to "promote understanding and debate around the objectives, intention and potential impact of data projects" and to "raise the profile of the importance of managing data ethically, along with a long-term goal of ensuring data ethics considerations are embedded in data projects".[48] It is intended to be used not only at the start of a project but also throughout its lifecycle.

The Canvas provides a framework, as opposed to a set of principles or a code of ethics, in order to aid in identifying and managing ethical issues. It can be used by individuals or discussion groups and workshops, ideally split into smaller groups, so as to collect ideas and inform the next steps to be taken.

The Data Ethics Canvas addresses issues and asks questions relating to:

- data sources, their limitations, and rights over their use;
- the sharing of data with other organisations;
- relevant legislation and policies;
- existing ethical frameworks;
- reasons for using the data;
- communication of the purpose of processing as well as risks and issues;
- positive and negative effects on people;
- minimising negative impacts;
- engaging with people;
- reviews, ethical monitoring, and iterations; and
- actions to take before moving on with the project.

[45] DataEthics, *Principles and Guidelines for Companies, Authorities & Organisations*, https://dataethics.eu/wp-content/uploads/Dataethics-uk.pdf
[46] https://theodi.org/article/data-ethics-canvas/
[47] ODI Data Ethics Canvas User Guide, https://docs.google.com/document/d/1MkvoAP86CwimbBD0dxySVCO0zeVOput_bu1A6kHV73M/edit#
[48] Ibid.

The conversation of these points should inform the ongoing management and discussion of data ethics, and to aid in the development of ethical frameworks and guidance tailored to the organisation.

2.5. Software development impact assessment

The Software Development Impact Statement, abbreviated SoDIS, describes a risk identification process that improves and expands risk perception. Enhanced risk perception is aimed at reducing the dangers of a narrow focus on quantitative risks and the number of software failures. The SoDIS process expands existing software development risk analysis methods by developers explicitly addressing a range of qualitative questions about the impacts of their development from a stakeholder perspective. By this, SoDIS overcomes limitations of traditional risk analysis such as the problem of addressing both quantitative and qualitative risks. SoDIS was tested successfully in organisations with different location, size, function, scope, development methodology, and technology level and against every phase of development.

The SoDIS process belongs to the family of issues-oriented approaches used in software systems development. It takes a comprehensive stakeholder perspective of the whole development cycle by considering each task within the structured plan of the project. A SoDIS risk analysis can be applied to any work product such as a work breakdown structure in a system's development.

The SoDIS process is a modification of the environmental impact statement process. It is used to identify potential negative impacts of a proposed project and specify actions that will mediate these anticipated impacts. It consists of four stages:

1. Identifying the project type together with immediate and extended stakeholders in a project
2. Identifying the tasks in a particular phase of a software development project
3. Associating every task with every stakeholder using structured questions to determine the possibility of specific project risks generated by that particular association
4. Completing the analysis by stating the concern and the severity of the risk to the project and the stakeholders, and recording a possible risk mitigation or risk avoidance strategy

The resulting document identifies all types of potential qualitative risks for all tasks and project stakeholders. The process is feasible as both bottom up and top down approach. It is very flexible, as it can be applied at any level of a hierarchy of tasks and any stage of the process can be revisited for any task level. According to Gotterbarn and Rogerson, SoDIS provides a pre-audit of the risk potential of the planned tasks before undertaking system implementation, which gives developers the opportunity to address the risks by mitigation or avoidance. The use of qualitative best practice questions associates a full range of stakeholders with the project tasks providing a comprehensive risk analysis which helps identify social, professional and ethical risks for a project. [49]

[49] Gotterbarn, D., & Rogerson, S. (2005). Responsible risk assessment with software development: creating the software development impact statement. Communications of the Association for Information Systems, 15(1), 40.

2.6. ISO/IEC 29134:2017 - Information technology -- Security techniques -- Guidelines for privacy impact assessment

A privacy impact assessment (PIA) is an instrument for assessing the potential impacts on privacy of a process, information system, programme, software module, device or other initiative which processes personally identifiable information (PII) and, in consultation with stakeholders, for taking actions as necessary in order to treat privacy risk. A PIA is a process that begins at the earliest possible stages of an initiative, when there are still opportunities to influence its outcome and thereby ensure privacy by design. It is a process that continues until, and even after, the project has been deployed.[50]

The ISO/IEC 29134:2017 provides guidelines for the process of conducting a PIA. The following steps are described in detail (preparation, performing the PIA, follow-up and report):

- Determine whether a PIA is necessary (threshold analysis)
- Preparation of the PIA
    - Set up the PIA team and provide it with direction
    - Prepare a PIA plan and determine the necessary resources for conducting the PIA
    - Describe what is being assessed
    - Stakeholder engagement
        - Identify stakeholders
        - Establish a consultation plan
        - Consult with stakeholders
- Perform the PIA
    - Identify information flows of PII
    - Analyse the implications of the use case
    - Determine the relevant privacy safeguarding requirements
    - Assess privacy risks
        - Privacy risk identification
        - Privacy risk analysis
        - Privacy risk evaluation
    - Prepare for treating privacy risks
        - Choose the privacy risk treatment options
        - Determine controls
        - Create privacy risk treatment plans
- Follow up the PIA
    - Prepare the report
    - Publication
    - Implement privacy risk treatment plans
    - Review and/or audit of the PIA
    - Reflect changes to the process

[50] ISO/IEC 29134:2017 Information technology -- Security techniques -- Guidelines for privacy impact assessment - https://www.iso.org/standard/62289.html

The ISO/IEC 29134:2017 provides detailed information for each step on the objective, input, expected output and action.

The risk criteria should reflect the organisation's values, objectives and resources. When defining risk criteria, the assessor should consider the following factors:

- Legal and regulatory factors that impact the safeguarding of the natural person's privacy and the protection of their PII
- External factors such as industry guidelines, professional and economic standards, company policies and customer agreements
- Factors predetermined by a specific application or in a specific use case context
- Other factors that can affect the design of information systems and the associated privacy safeguarding requirements

2.7. e-SIDES Deliverable 2.2 Applied Ethics Toolkit

Within Deliverable 2.2 of the E-Sides project we have identified ethical values to consider for big data technologies and legal, societal and economic issues that regularly emerge within different big data contexts. Deliverable 2.2 also developed an Applied Ethical Toolkit for practical use. We consider this toolkit as useful to provide practical assessment criteria for those professionals that are active in the field of big data technologies, in order to actively assess and implement the ethical considerations by offering recommendations according to four phases of big data development and application. The toolkit is meant for professionals to use it before a big data project starts as well as throughout the employment of these technologies, in order to ensure that these ethical issues are properly addressed.

Mepham's pluralism of principles[51]	***Technomoral virtues***[52]	***Values from value-sensitive design(VSD)***[53]	***Values from Anticipatory technology ethics*** [54]	***Values in biomedical ethics***[55]	***e-SIDES: values for big data technologies***
Care for well-being	**Care**	**Human Welfare**	**Well-being and the common good**	**Beneficence**	**Human welfare**
	Magnanimity, Courage	**Autonomy**	**Autonomy**	**Autonomy**	**Autonomy**
	Humility, Self-control	**Calmness**	**Health, (no) bodily and psychological harm**	**Non-maleficence**	**Non-maleficence**
Respect for justice	**Justice**	**Freedom from Bias; Universal usability**	**Justice (distributive)**	**Justice**	**Justice (incl. equality, non-discrimination, digital inclusion)**

[51] Mepham, B. (2010) 'The Ethical Matrix as a Tool in Policy Interventions: The Obesity Crisis', in (F-T. Gottwald et al. eds) Food Ethics, Springer Science Business Media, pp. 17-28
[52] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, pp. 120-121
[53] Friedman, B. et al. (2006) 'Value Sensitive Design and Information Systems' in (Zhang, N. P. and Galletta, D. eds.) Human-Computer Interaction in Management Information Systems: Foundations, M.E. Sharpe Publishing, pp. 19
[54] Brey, P. (2012) Anticipatory Ethics for Emerging Technologies, Nanoethics 6(1), 1-13
[55] Beauchamp, T. and Childress, J. (2012) Principles of Biomedical Ethics, 7th edition, New York, Oxford University Press, pp.

	Perspective	**Accountability**	**N/A**	**N/A**	**Accountability (incl. transparency)**
	Honesty, Self-control	**Trust**	**N/A**	**Veracity**	**Trustworthiness (including honesty and underpinning also security)**
Respect for dignity	**N/A**	**Privacy; Informed Consent; Ownership and Property**	**Rights and freedoms, including Property**	**N/A**	**Privacy**
	Identity	**Identity**	**Human dignity**	**Respect for dignity**	**Dignity**
	Empathy, Flexibility, Courage, Civility	**Courtesy**	**N/A**	**N/A**	**Solidarity**
	Courage, Empathy	**Environmental Sustainability**	**(No) environmental harm, Animal welfare**	**N/A**	**Environmental welfare**

Table 1: Virtues to uphold during techno-social change in regard to big data technologies

Step One: "To identify which ethical values are relevant to the big data practice in question."

The e-SIDES project identified human welfare, autonomy, non-maleficence, justice, accountability, trustworthiness, privacy, dignity, solidarity, and environmental welfare as being particularly relevant to big data technologies. Yet, not all of these ethical values may be applicable to a given big data application area. In this phase the given big data technology shall be carefully analysed first.

Step Two: "Recognise the options for response."

During this phase "decision makers should consider all available options to remedy the ethical conflict." Different arguments shall be considered in light of ethical values which might be at stake when options for a certain course of action are proposed. Each option shall be critically assessed. Furthermore, such evaluation shall not merely mean a 'box to check' throughout the process, but should be regarded as being an essential step to take in between stages.

Step Three: "Recommend and implement a response."

In this phase, it should be assessed: "Will the proposed response lead to conflicts with other ethical values? If so, will appropriate measures be implemented to address those concerns? Will the proposed response uphold the most ethical values, and to the greatest extent, in comparison to other proposed responses?" In this stage these questions need addressed while bearing in mind future ethical challenges which may arise as a consequences of certain choices during this phase.

Step Four: "Anticipate further ethical conflicts (and repeat)."

"Is the organisation planning to add new features or to improve upon functionality in the big data technology? Is the organisation planning to utilise new, different, or more complete data sets in the technology? Is the organisation planning to use algorithms or machine learning in ways they were not previously used?" If the answer is yes to these questions, then the ethical issues need to be re-evaluated. Such re-evaluation does not necessarily have to be a surprise, but rather more a consequences of interrelated processes and stages during big data developments.

3. Assessing impacts of introducing data-driven innovation and privacy-preserving technologies in different big data contexts

In this section, through literature review, first we examine data-driven innovation and privacy-preserving technologies in three big data contexts: healthcare, transportation and smart cities, and web browsing and third-party tracking. Second, based on desk research and our conducted interviews, we show the extent to which impact assessment tools and mechanisms are used within those Horizon 2020 data related projects as detailed in the introduction. We explore this in order

3.1. Healthcare

With the significant amount of sensitive data at play, healthcare is an area where providing specific protection to special categories of data is not only legally required[56], but is essential also from an ethical point of view. Under Article 9 of the GDPR, 'the lawfulness of processing special categories of data', healthcare information counts as a special category of personal data and requires advanced protection. There is a significant amount of research assessing the implications of privacy-preserving and security technologies around devices for healthcare.[57] Academics on healthcare data[58] have demonstrated that assessing the impact can be context-dependent and device-dependent as the ways in which data is collected and processed across technological tools differs per healthcare application.[59] For instance, a widely used privacy-preserving technology in healthcare is homomorphic encryption. Certain researchers argue that complex algorithms must be disintegrated into simpler operations; a framework using homomorphic encryption, for instance, must be disintegrated in terms of simple homomorphic additions and multiplications which can also be considered as very useful safeguards for successful societal, ethical and legal assessments.[60] Furthermore, researchers focusing on different types of healthcare devices also highlight that "addressing privacy concerns requires addressing security issues like access control, authentication, non-repudiation, and accountability, without which end-to-end privacy cannot be ensured."[61] Privacy-preserving solutions in healthcare also include blockchain

[56] Article 4(13), (14) and (15) and Article 9 and Recitals (51) to (56) of the General Data Protection Regulation
[57] Iwaya, L. H.; Martucci, L. A.; Fischer-Hubner, S. (2016) - Towards a Privacy Impact Assessment Template for Mobile Health Data Collection Systems, https://www.diva-portal.org/smash/get/diva2:1043663/FULLTEXT01.pdf
[58] Adams, S. & Van Veghel, D. & Dekker, L. (2015). Developing a Research Agenda on Ethical Issues Related to Using Social Media in Healthcare; Cambridge Quarterly of Healthcare Ethics:CQ The International Journal of Healthcare Ethics Committees. 24. Pp. 293-302. 10.1017/S0963180114000619.
[59] Kim J. W., Jang B., Yoo H. (2018) Privacy-preserving aggregation of personal health data streams. PLoS ONE 13(11): e0207639. https://doi.org/10.1371/journal.pone.0207639, Retrieved on 10th of May 2019 from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207639
[60] Sharma, S.; Chen, K.; en Seth, A. (2018) Towards Practical Privacy-Preserving Analytics For IoT and Cloud-Based Healthcare Systems, IEEE Internet Computing, March-April 2018. https://arxiv.org/ftp/arxiv/papers/1804/1804.04250.pdf
[61] Sahi, M. A.; Abbas, H. et al. (2018) Privacy-preservation in e-Healthcare Environments: State of the Art and Future Directions, IEEE Access Special Section on Security Analytics and Intelligence for Cyber Physical Systems - https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8089328

technologies. However, the potential of these technologies is not yet fully exploited according to researchers due to the fact that most investors are unfamiliar with this technology and lag behind in creating a sufficiently user-friendly interface and in educating users.[62] Yet, in their assessment the potential of blockchain technologies for healthcare lies in the fact that it is a distributed ledger and can maintain the high security of patients' data and by this meet not only GDPR requirements, but potentially also criteria for ethical, legal, social and economic impact assessments.

Within the context of healthcare, the Horizon 2020 projects called BODYPASS, MyHealthMyData, CLARUS and BigMedilytics were assessed in terms of the privacy-preserving technologies they used and the kinds of assessments tools they considered for evaluating the impact of their technologies. We can observe in general the followings about these projects:

- Each of them deals with sensitive data and the purposes of processing can differ significantly. The projects implement different privacy-preserving technologies to reach these purposes, while not all of them have impact assessment tools. Some projects have ethical committees they gain advice from.
- Given in healthcare it is sensitive data that is processed for different purposes such as, for instance, research, medical examination or prognosis on diseases we have observed that certain projects tend to integrate multiple privacy-preserving technologies in a combined manner in order to capitalize upon the different features of these technologies in relation to different stakeholders. For instance, when it is not necessary to know the personal details attached to a disease the anonymization tool is considered useful and statistical research from such data can still be conducted. Access control and the sharing of data were also underlined by the projects: for the sharing of sensitive data only with those that are authorised block chain was embraced.
- Healthcare projects are highly aware of their duty to comply with the GDPR, some stakeholders, such as hospitals seemed extra careful in sharing data even within secured environments due this awareness. Applicable corporate policies are usually taken into account by formulating system requirements accordingly.

3.1.1. BodyPass

The first project we assess here is Bodypass.[63] It is a Horizon 2020 project and part of the Big Data Value Public-Private Partnership. BodyPass focuses on data-driven 3D scanning and reconstruction (D3DR) technologies. It started in January 2018 and will last until the end of 2020.[64] The project (API-ecosystem for cross-sectorial exchange of 3D personal data) aims to break barriers between the health sector and the consumer goods sector, and to eliminate current data silos. The main objective of BodyPass is to foster exchange, linking and re-use, as well as to integrate 3D data assets from the two sectors. For this,

[62] Vazirani AA, O'Donoghue O, Brindley D, Meinert E (2019) - Implementing Blockchains for Efficient Health Care: Systematic Review J Med Internet Res Vol. 21 No. 2: e12439 DOI: 10.2196/12439
[63] http://www.bodypass.eu/
[64] Bodypass project - https://cordis.europa.eu/project/rcn/212483/factsheet/en

BodyPass has to adapt and create tools that allow a secure exchange of information between data owners, companies and subjects (patients and customers). The project looks to develop tools to access huge data sets from the health sector and consumer goods industries in order to extract useful 3D data information for medical applications and product design.

As there are no deliverables apart from the dissemination plan available, it is difficult to evaluate whether ethical and societal aspects are addressed. There is, however, a work package on the validation in real contexts and one on ethics requirements.[65] According to its brochure, BodyPass allows a secure exchange of information and privacy-preserving methods for exchange of 3D images and aggregated data.[66] Furthermore, the results of the conducted assessments by BodyPass show that 2D3D scanners are as reliable and privacy-preserving as high-resolution 3D scanners.

The consortium members of the BodyPass project explained that they do not have any tools for assessing the impact of their big data solution on society during the project.

The consortium members also explained that they integrate privacy-preserving technologies into their big data solutions (i.e., security and privacy features), by stating that they would develop specific anonymisation tools for 3D scanners. Furthermore, they added that because the scans can include face and other personal data the anonymisation tool and process is tailored for these types of data.

The BodyPass project partners also explained that they take preventive measures to avoid data breaches by the followings: BodyPass is based on application programming interfaces (APIs) and access to data is granted by such APIs and each data provider must assure their own security.

In terms of ensuring compliance with laws and corporate policies, they answered that they indeed have done a data protection impact assessment, but have not explained further how they have done the assessment.

The Bodypass project also highlighted that all their work is reviewed by an ethical committee, but have not given further details as to what that assessment looks like.

With respect to finding a balance between protecting economic interests and avoiding undesired ethical and societal implications. The BodyPass consortium member answered that they dealt with data from hospitals and hospitals were very reluctant to share healthcare data even if it was allowed to share data by the GDPR. This was so because hospitals were afraid of bad publicity.

3.1.2. MyHealthMyData

MyHealthMyData (MHMD) is a Horizon 2020 ICT-2018-16 Research and Innovation Action project. It started in November 2016 and ends in December 2019. MyHealthMyData "*aims at fundamentally changing the way sensitive data are shared. MHMD is poised to be the first open biomedical information*

[65] http://www.bodypass.eu/content/workpackages
[66] www.bodypass.eu/sites/default/files/bodypass/public/content-files/article/BodyPass_brochure_online_0.pdf

network centred on the connection between organisations and individuals, encouraging hospitals to start making anonymised data available for open research, while prompting citizens to become the ultimate owners and controllers of their health data. MHMD is intended to become a true information marketplace, based on new mechanisms of trust and direct, value-based relationships between EU citizens, hospitals, research centres and businesses." [67] Furthermore, MyHealthMyData aims to address the problem that stems from the fact that hospitals are regularly large, isolated data repositories and often technologically not equipped with sharing patients' data in a responsible manner.

We have not interviewed the consortium members of MyHealthMyData. Nevertheless, the following can be explained about the project in terms of impact assessments: With respect to whether MyHealthMyData assesses the impact of their big data solution on society we have found that MyHealthMyData, for instance, uses an inventory of possible security threats in order to assess and minimize the risks. Furthermore, MyHealthMyData develops its own security infrastructure that considers three different factors: "*(i) the current/existing version of the MHMD platform, its components and their interactions, (ii) state-of-the-art solutions in the security context, by evaluating the [security] aspects …, and (iii) the possible consideration of security and privacy related standards and regulations. Combining these factors, we are able to provide a security infrastructure capable of considering different contexts, focusing on the aspects that are critical for the MHMD platform, and also evaluating aspects for exploitation and scalability, through standardization procedures and regulations. Also, the security of the entire MHMD platform is improved thanks to the MHMD distributed Intrusion Detection System (MHMDdIDS), an innovative Intrusion Detection System aimed to monitor and protect the entire system.*"

As to the extent that MyHealthMyData integrates privacy-preserving technologies into their big data solutions (i.e., security and privacy features), we have found that MyHealthMyData offers possibilities to exploit and integrate privacy-preserving technologies into big data solutions such as dynamic consent, blockchain-based software infrastructure, personal data accounts, and smart contracts, in addition to being able to situate these technologies within a legal and regulatory framework that facilitates anonymised data production for open research.[68]

From the MyHealthMyData project deliverables we could find that MyHealthMyData is heavily focused on strong security measures. Beyond the above-mentioned privacy-preserving technologies, MyHealthMyData also has a tool called MHMD Driver that is aimed at allowing for access to existing consent for each data item. "*By mapping such information with the study definition, it is possible to filter out the data items that do not match the consent, hence preventing leakage of data without proper consent.*"[69]

[67] Blockchain to enable medical data to be stored and transmitted safely and effectively - https://ec.europa.eu/digital-single-market/en/news/blockchain-enable-medical-data-be-stored-and-transmitted-safely-and-effectively

[68] Cambiaso, E.; Vaccari, I.; Aiello, M. - Deliverable 5.4, http://www.myhealthmydata.eu/deliverables/D5.4-security-infrastructure.pdf

[69] Ibid.

3.1.3. CLARUS

The CLARUS Project is a Horizon 2020 project started in January 2015 and ended in December 2017. Although this project has already ended, we included it here as lessons that can be learned from the ways in which it used assessment tools for privacy-preserving technologies. CLARUS was described as "*aligning the interests of cloud service providers and their customers for no-compromise security of all data stored in the cloud*".[70] The CLARUS project uses five types of privacy-preserving technologies: data anonymisation, data encryption, data splitting, homomorphic encryption, and searchable encryption.

The **CLARUS** (A Framework for User Centred Privacy and Security in the Cloud) project[71] aimed to enhance trust in cloud computing services by developing a secure framework for the storage and processing of data in the cloud that allows end users to monitor, audit and retain control of the stored data without impairing the functionality and cost-saving benefits of cloud services. CLARUS provides the end user with a dedicated proxy located in a trusted domain implementing security and privacy features towards the cloud provider. CLARUS also provides a set of security auditing services enabling the user to supervise the security operations performed by the CLARUS framework as well as other trust-enhancing features. These services include a Security Policy Manager, an Access Rights Management tool and a Security Administrator.

The CLARUS project uses several privacy techniques for the proxy and the tools, which include data anonymisation, data encryption, data splitting, homomorphic encryption and searchable encryption.[72] The project also developed some legal guidelines for data quality and data subject rights. The data quality principles consist of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality and the data subject's rights consist of information, access, rectification, erasure, restriction and portability. Furthermore, CLARUS published a deliverable on the legal and ethical requirements of the project.[73] The requirements range from the above-mentioned guidelines to general security guidelines, breach notification and contractual mitigation of liability. All the identified requirements are considered a must have, meaning "the project will be seriously impacted if the requirement is not net."[74]

The CLARUS project had different case studies, and one of them is eHealth. Within the eHealth case study, the CLARUS project focused on the management of healthcare records. Patient records contain highly sensitive and personally-identifiable data and therefore they are securely stored in the CLARUS cloud. Privacy-preserving mechanisms "*shall be implemented before outsourcing obfuscated data to the cloud with preservation of functionalities*". Given that patient records also contain heterogeneous data

[70] http://clarussecure.eu/
[71] http://clarussecure.eu/
[72] http://clarussecure.eu/privacy-techniques
[73] D2.4 – Legal and Ethical Requirements, 2017-02-17, http://clarussecure.eu/sites/default/files/CLARUS-D2.4-LegalandEthicalRequirements.pdf
[74] ebd. p. 58

in terms of "*structure and type CLARUS must demonstrate high level of data concealment in all situations*".[75]

With respect to the interview questions, we have received the following answers. On the question as to whether the CLARUS project assessed the impact of their privacy-preserving technology on society, they answered that they tested the impact through two CLARUS case studies. One case study: "*Storage and processing of privacy-protected medical records in the cloud. This was carried out by Hospital Clínic of Barcelona, which was a CLARUS partner.*" Second case study: "*Storage and processing of geolocated data in the cloud. This was carried out by AKKA, a French company that was also a CLARUS partners. After the end of the project, AKKA continues to use the CLARUS product.*" On the question whether CLARUS integrates privacy-preserving technologies (i.e., security and privacy features) into their big data solution, "*CLARUS was precisely about protecting sensitive data so that they cannot only be stored in the (untrusted) cloud, but also processed using the cloud's computing power. Three categories of privacy-preserving technologies were used: anonymization, splitting and searchable encryption.*" On the question as to whether CLARUS took preventative measures to avoid data breaches CLARUS answered that: "*all data stored in the cloud were privacy-protected using the aforementioned technologies*".

On the question as to whether they took measures to ensure compliance with laws and corporate policies, CLARUS consortium members answered the following: "*The aim of CLARUS was to reconcile GDPR compliance with using untrusted clouds for storage and processing of sensitive data.*" On the question as to whether they adopted any best practices to ensure that undesired ethical and social implications of their big data solution are recognised and addressed, they stated that *"Developing and deploying privacy-preserving technologies has precisely the aim of avoiding undesired ethical and societal implications."*

On the question as to whether they adopted any best practices to ensure that undesired ethical and societal implications of your big data solutions are recognized and addressed, they answered: "*Developing and deploying privacy-preserving technologies has precisely the aim of avoiding undesired ethical and societal implications.*" On the question whether they considered to find a balance between protecting economic interests and avoiding ethical and societal implications and in what respect, they answered that: "*There is a trade-off. For example, the CLARUS solution places constraints on the type of processing that the cloud can perform on the privacy-protected data. E.g. if you encrypt sensitive data with searchable encryption, the encrypted data can only be searched. If you anonymize data, then the cloud computations yield results that are not exact.*"

3.1.4. BigMedilytics

The BigMedilytics (Big Data for Medical Analytics)[76] project aims to enhance patient outcomes and increase productivity in the health sector by applying big data technologies to complex datasets, while ensuring security and privacy of personal data. The 38-month project started in January 2018 and will end in February 2021.

[75] Privacy-preserving techniques for no-compromise security in the cloud - http://www.clarussecure.eu/sites/default/files/ClarusWhitePaper_B5_Web_0.pdf
[76] https://www.bigmedilytics.eu/

The project implements twelve pilots that cover three themes that have the potential to have a large impact on the sector. The themes are population health and chronic disease management, oncology and industrialization of healthcare services.

Several general deliverables have been released so far.[77] D1.2, among others, details the project's activities regarding security and privacy of data access and processing.[78] BigMedilytics has a dedicated task that focuses on questions related to security and privacy. Moreover, the project has established an ethics board that is headed by a data protection officer (DPO). The DPO advises the consortium on how ethics-related matters should be handled. In addition, the project is also advised by an independent, external ethics advisor.

Big Data Healthcare Analytics Blueprint and the the BigMedilytics-BigMatrix are aimed at mapping requirements and the technical components for data sharing. The Matrix is multidimensional as it takes aspects of technologies, of pilots, businesses, communities and aspects of specific data sources. The matrix also distinguishes between "*data sources that have different velocities, e.g. mobile devices vs. sensor streams for telemedicine pilots or real-time location data vs. electronic medical records for hospital workflow focused pilots in the Industrialization of Healthcare theme.*" BigMedilytics-BigMatrix provides an overview of security-preserving data access, processing and access control with support for auditing. The BigMedilytics employs technical methods for security preservation and de-identification of electronic medical records (EMR), real-time-locating systems (RTLS) and the use of HTTPS protocols and a minimum security feature: single-factor authentication. The project also uses virtual private networks (VPN) access for external users including two-factor authentication, and access to the administrative node for partners provided by the hypervisor of the cluster of technologies.[79]

The combination of these privacy-preserving technologies facilitate auditing mechanisms and through this they aim to avoid the detrimental societal, legal and ethical impact of using these technologies. Through the combination of technologies and data sharing agreements, BigMedilytics aims to manage negative effects by taking aspects of technologies, of pilots, businesses, communities and aspects of specific data sources into account. While taking these aspects into account they could rely upon the Data ethics Impact Assessment and Guidelines, the Data Ethicas Canvas and e-SIDES Applied Ethics Toolkit.

3.1.5. Recommendations

The four assessed projects demonstrate that the field of healthcare is scattered into diverse sub-fields with varying sub-sectoral goals and sometimes diverging interests. Furthermore, it also showed that the same privacy-preserving technologies can be employed for different healthcare purposes involving data, the impact assessments of which are sometimes not taken into account during the design. For instance,

[77] https://www.bigmedilytics.eu/deliverables/
[78] https://www.bigmedilytics.eu/wp-content/uploads/2018/12/D1.2_Initial-prototypes_v1.0.pdf
[79] Deliverable 1.2 - Initial prototypes for specific components for all Big Medilytics pilots - https://www.bigmedilytics.eu/wp-content/uploads/2018/12/D1.2_Initial-prototypes_v1.0.pdf

BodyPass provides application programming interfaces but they do not employ assessment tools during the project.

They could capitalise on employing the guidelines of the Art. 29 Working Party, the Data Ethics Canvas and also the Applied Ethics Toolkit from our Deliverable 2.2. The Art. 29 WP Guidelines can help to evaluate the extent to which the sharing of body scans is proportionate and necessary. Furthermore, the Applied Ethics Toolkit could aid in determining how is accountability organised among developers and customers.

The MyHealthMyData project encompasses a security infrastructure that is capable of considering different contexts and different big data and healthcare-related aspects. Within MHMD, blockchain technologies are used that enable medical data to be stored and transmitted safely and effectively. MHMD could benefit from using the Data Ethics Impact assessment tool as they took effort in becoming GDPR-compliant. The CLARUS project has already ended, but on the basis of their shared answers on the survey they employed three privacy-preserving technologies: *anonymisation*, *splitting* and *searchable encryption.* Yet, the employment of these three technologies does not say much about the underlying processes. Hence, the Article 29 Working Party Guidelines, the Data Ethics Impact Assessment guidelines, the Data Ethics Canvas could be beneficial to integrate into their processes in order to assess ethical, legal and social impacts.

The BigMedilytics project, among the four assessed projects, appeared to be the most comprehensive in terms of aligning the interests of different communities, businesses and aspects of technologies and data. Therefore, in that alignment work the use of The United Nations Global Pulse Data Innovation Risk Assessment Tool, the Data Ethics Canvas and the Applied Ethics Toolkit of the e-SIDES project could be of further help.

3.2. Transportation

Transportation and smart cities are characterised by increasing ubiquity and combinability of digital and autonomous technologies. They allow for the exponential lengthening of the big data lifecycle. Given the accumulation of purposes for which big data is used in transportation and smart cities, such as urban safety, environmental friendliness and more efficient transportation and city management, all of these purposes are increasingly aimed to be achieved by forecasting and prediction that are based on the personalised experiences of city inhabitants.

The following observations were made with respect to projects focusing on transportation and smart cities:

- Projects are aware of the impacts of the big data solutions they develop on the society. They assess these impacts but usually not in a highly systematic way or following a rigorous methodology. Some established ethics advisory boards.
- Projects tend to integrate common privacy-preserving technologies into their solutions. If possible, personal data is anonymised or at least processed, stored and transmitted in ways that are secure, for instance, by using encryption. Policy enforcement and user control also receive considerable attention.

- Projects generally stress that they are compliant with applicable laws. The GDPR or its predecessors or national privacy laws are frequently mentioned. Applicable corporate policies are usually taken into account by formulating system requirements accordingly.
- The extent of attention projects pay to societal and ethical issues varies greatly. If projects take measures to assess impacts and mitigate them, they usually implement preventive measures. Reactive measures do not seem to play a significant role.
- The alignment of people, processes and technology is typically not made explicit. However, project documentation clearly shows that projects acknowledge that people and processes are at least as important as technology.
- The adoption of best practices is very common among projects. Projects build upon lessons learned and implement whatever has proven useful and does not interfere with the project's specific objectives.
- Projects focusing transportation and smart cities do generally not see a conflict between protecting economic interests and avoiding undesired ethical and societal implications. On the one hand, economic interests are not overvalued and, on the other hand, undesired ethical and societal implications than cannot be mitigated are considered as show-stoppers.
- The measures taken in the transportation and smart city context do not seem to differ much from those application contexts. Personal data is not essential for many big data scenarios related to transportation and smart cities.

3.2.1. SPECIAL

The SPECIAL[80] project[81] addresses the contradiction between big data innovation and data protection by proposing a technical solution that makes achieving both of these goals more realistic. The solution is an extensible environment for managing personal data usage policies, ensuring compliance with such policies and tracking personal data usage along with the contexts it is being used in.
The project focuses on three pilots.[82] One of them, which focuses on data collected by T-Mobile Poland, is relevant from a transportation perspective. T-Mobile Poland collects raw data (i.e., information about the calls, who calls from where for how long etc.) from its base stations. The number of data points (unique pieces of information) totals to 45 billion. The data is stored in Hadoop databases in a manner that supports analytics and intelligence gathering. Such data intelligence could be used for individual location-insight services, aggregated location-insight services, individual location-based services and aggregated location-based services. In order to exemplify the range of services and the value that could be offered, particular attention is paid to three distinct yet related scenarios:

Within the scope of the **municipality road layout** scenario the sharing of location data with local governments in order to optimise public infrastructure is investigated. T-Mobile Poland would like to

[80] https://www.specialprivacy.eu/
[81] There has been a quite intense cooperation between SPECIAL and e-SIDES.
[82] SPECIAL D1.5 Use case scenarios V2 - <https://www.specialprivacy.eu/images/documents/SPECIAL_D15_M14_V10.pdf>

give information to local governments concerning how often people commute to which regions, what paths they are taking etc., in order to lay out local roads in the optimal way. This requires linking age, usage, time of day and geolocation information and sharing it with a third party.

The **bank travel insurance** scenario involves the combination of location and banking data to offer cheaper travel insurance clients. A Bank would like to offer better travel insurance to clients that travel often. They share their phone number database with T-Mobile Poland and would like to know which of these phone numbers often use roaming. This scenario could be problematic from a sharing perspective as both T-Mobile Poland and the bank will need to get the relevant consent for both processing and sharing.

Providing real-time traffic alerts to clients is the aim of the **traffic condition warning scenario**. T-Mobile Poland would like to send out push information suggesting that users avoid particularly busy roads or public transportation lines based on their location.

A project representative kindly provided answers to the key questions that guided our analysis. The representative highlighted that although the main goal of SPECIAL is to help companies to continue to innovate in light of the GDPR, the project is guided by the need to support not only companies but also data subjects (in terms of enabling them to control how their personal data is used data) and regulators (in terms of supporting them in ensuring that companies comply with the GDPR).

With respect to ethical and societal impacts and their assessment, it was emphasised that SPECIAL aims to allow citizens and organisations to share more data, while guaranteeing data protection and compliance, thus enabling the creation of new value and insight from shared data. Moreover, the representative pointed out that the SPECIAL vision is realised and validated via real world use cases that need to process and share data in a privacy-preserving manner in order to deliver improved services.

Privacy-preserving technologies play a key role in the context of the SPECIAL. From a privacy perspective, SPECIAL proposes a scalable policy management framework, which supports policy specification and administration, needed in order to give users control of their personal data and to ensure that both access/usage policies and legislative requirements can be represented in a machine-readable format, which can be verified automatically by their compliance checking algorithms. From a security perspective, the SPECIAL transparency and compliance framework is used to verify that data processing and sharing events comply with the data usage policies specified by the data subject. A combination of encryption, hashing and digital signatures is used to ensure both the integrity and nonrepudiation of policies and events.

The project takes data breaches seriously and takes preventive measures to avoid them. SPECIAL enables data controllers to evaluate and better control the risk of law infringements and data breaches because the transparency and compliance platform allows for a much better assessment of processes and resulting risks. In keeping with the requirements of the GDPR, the SPECIAL dashboard supports automatic breach notification providing information concerning non-compliance to data protection officers and data subjects in a non-intrusive manner.

With respect to ensuring compliance with laws and corporate policies, the representative stated that the SPECIAL policy language can be used to specify usage policies, regulatory requirements and business

policies. The SPECIAL policy language, vocabularies and compliance checking algorithms enable companies to develop and run their data-driven business by helping them to comply not only with the data subject's wishes but also with data protection legislation.

Both legal and ethical considerations are, according to the project representative, core to the success of SPECIAL. The use case partners are already extremely conscious of the sensitivity of the data that they hold and currently, when it comes to processing personal data, they are on the side of caution. The legal and ethical objectives of SPECIAL are twofold, on the one, hand the project aims to provide a guarantee to companies that they are abiding by both the legislation and usage policies defined by data subjects and, on the other hand, SPECIAL will provides much more visibility to data subjects concerning the processing of their personal data.

The representative is confident that if the use cases within the SPECIAL project work, this may trigger a new strategy of organisations doing business in Europe. This strategy could be to embrace data protection-friendly approaches more in the future as something that can also be a competitive advantage. Therein, the project's approach to support informed consent acquisition easily manageable by the organisation as well as by the data subjects themselves, may contribute to a reconciliation of both economic interests and informational self-determination in the digital era.

3.2.2. AEGIS

The AEGIS[83] project creates an interlinked "Public Safety and Personal Security" (PSPS) data value chain, and delivers a novel platform for big data curation, integration, analysis and intelligence sharing. Through the AEGIS platform, the PSPS data value chain analysis is conducted at multiple levels: (I) data privacy enhancement, (II) data pre-processing, (III) big data analysis, and (IV) data intelligence sharing.[84] Moreover, the project develops an automotive and road safety demonstrator, a smart home and assisted living demonstrator and an insurance sector demonstrator.[85] From a transportation perspective, the road safety demonstrator is most interesting.

The road safety demonstrator explores how vehicle driving data and other road safety-related data can be meshed and modelled, aggregated and semantically annotated in order to extract meaningful, automotive and road safety-relevant information for various stakeholders.[86] The demonstrator is developed according to three different scenarios:

The **broken road indicator** provides insights into road conditions based on exploiting individual vehicle sensor data, traffic data and map data.

The **safe driving indicator** infers the driver's safety style and then calculate a safety index, through utilising vehicle sensor data along with environmental information and other content.

[83] https://www.aegis-bigdata.eu/
[84] https://www.aegis-bigdata.eu/approach/
[85] https://www.aegis-bigdata.eu/demonstrators/
[86] https://www.aegis-bigdata.eu/demonstrators/automotive-and-road-safety-demonstrator/

The **regional driving style risk estimator** calculates a regional driving safety risk metric for certain regions including intersections, streets, cities or countries.

The demonstrator architecture is illustrated in Figure 1. The core module, the AEGIS platform, is responsible for the aggregation of data from various open sources as well as from platform end users. Capabilities for conducing data science and data analysis are continuously demanded by the mobility domain. The AEGIS platform plays a fundamental role as it provides the required data analysis infrastructure for the data scientist at Virtual Vehicle (VIF). VIF[87] is an international research and development centre for the automotive and rail industries and a partner of the AEGIS project.

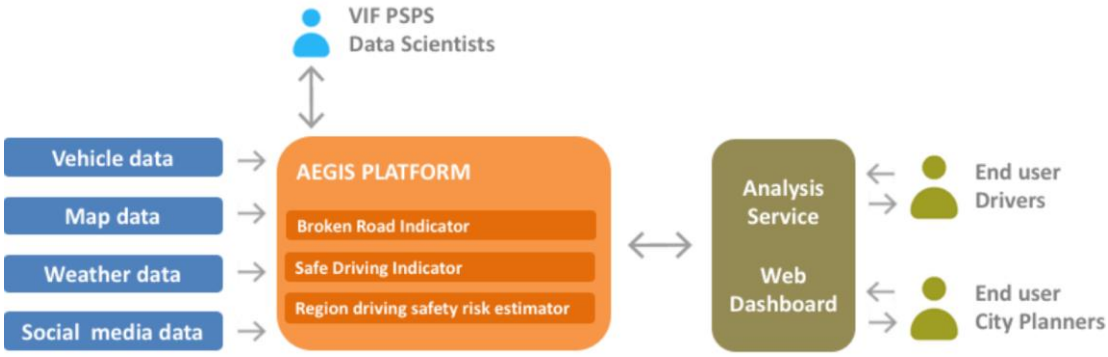

Figure 1: The road safety demonstrator's architecture

Consideration of ethical and societal aspects is an important part of the AEGIS project. It has an own ethics advisory board (EAB) working closely with the AEGIS consortium.

The EAB assessed each demonstrator comprehensively.[88] The road safety demonstrator is described as not linked with self-driving car technology. Thus, ethical problems of decision making by artificial intelligence does not arise. However, it is concerned with collecting data from vehicles operated in the field by volunteering drivers (who sign an informed consent). The EAB emphasises that all (vehicle operation) data processed is anonymised. According to the EAB, one ethical issue could come up, if the information which would be delivered by the demonstrator would be invalid and not reliable.[89] If this is the case, there is 1) an empirical question which cannot be addressed yet and 2) an ethical question: which degree of validity and reliability will be "sufficient" and will not raise ethical questions? This question must be and can be discussed in course of time.

Data protection problems of the demonstrator are of minor importance, according to the EAB. Data to be collected during the experiments is sensor data (e.g., speed, acceleration, etc.) and/or simulation

[87] https://www.v2c2.at/

[88] D9.3 Ethics Advisory Board's Report, https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D9.3-GEN-Requirement-N-%C2%B0-4-_EABs-Report_v1.0_public.pdf

[89] See also Custers, B.H.M. (2003) Effects of Unreliable Group Profiling by Means of Data Mining. In: G. Grieser, Y. Tanaka and A. Yamamoto (eds.) Lecture Notes in Artificial Intelligence, Proceedings of the 6th International Conference on Discovery Science (DS 2003) Sapporo, Japan. Berlin, Heidelberg, New York: Springer-Verlag, Vol. 2843, p. 290-295.

data. Sensor data is generated through connecting a device developed by one of the project partners to the onboard diagnostic (OBD2) interface of a car. Simulation data is generated by study participants using a driving simulator and may include many additional values. Both sensor data and simulation data have to be stored on a research server to allow the development of algorithms for inferring events including broken roads, patterns of safe and unsafe driving or driving risks. Sensor and simulation data are kept on this server till the end of the project. Of course, all data is anonymised before being stored on the server. The EAB stresses that the road safety demonstrator does not involve processing any personal data. However, according to the corresponding business scenarios and business models developed in the project and aiming to scale these applications to the market, a future collection of personal data might be considered. The EAB concludes that there are currently no issues of data protection and ethics that have to be checked. Moreover, it is stated that suitable mechanisms are in place to deal with all ethical risks, including data protection that can potentially arise during the course of the project.

The EAB ensures strict adherence to privacy-by-design and privacy-by-default paradigms. Moreover, the EAB points out that all partners are aware of their responsibilities and obligations to respect the data subjects' privacy, confidentiality but also autonomy, self-determination and dignity, in compliance with European standards and best practices. Apart from the establishment of an EAB and the adherence to privacy-by-design and privacy-by-default paradigms, AEGIS elaborated a specific ethical, privacy and data protection strategy at the beginning of the project, pays attention to legal compliance and put in place proper informed consent procedures for AEGIS recruitment.

The demonstrators were also assessed by the project consortium paying particular attention to compliance with data protection and ethical awareness.[90] With respect to compliance with data protection, data protection principles, data processing, and technical and organisational measures were considered. While the principles included data minimisation, purpose limitation, storage retention, data accuracy, integrity and confidentiality, and fair data processing, with respect to data processing, lawfulness, compliance with obligations and respect for data subject rights were addressed. Privacy-by-design, anonymisation and privacy-by-default were relevant with respect to measures. With respect to ethical awareness, administrative requirements such as national legislation, confidentiality and access restriction and voluntariness of participation received particular attention.

3.2.3. Transforming Transportation

The Transforming Transportation (TT)[91] project[92] demonstrates the transformations that big data brings to the mobility and logistics market in terms of, for instance, increased operational efficiency, improved

[90] D9.1 Other ethics issues, https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D9.1-Other-Ethical-Issues_v1.0-public.pdf
[91] https://transformingtransport.eu/
[92] The TT project has cooperated with e-SIDES in several respects.

customer experience and new business models. The project focuses on how big data reshapes transport processes and services using pilots from seven domains that are particularly important for the sector: [93]

Smart highways cater for a number of different activities to incorporate technologies into roads for improving the operation of autonomous cars, for lighting, for monitoring the condition of the road, targeting improved traffic distribution, reduced accidents, better security, reduced operational costs, enhanced optimisation of resources for road operators, as well as data forecasts and applications for users and operators.

Sustainable vehicle fleets are outfitted with special technologies that tap into the Internet or wireless LAN and provide additional benefits to drivers. Through such technologies estimated savings in operating expenses are achieved through better maintenance, reduction in fuel consumption based on better routing and driving patterns, enhanced safety, predictive maintenance through pattern recognition, as well as enhanced and more competitive insurance models.

Proactive rail infrastructures invest in cutting edge technology enabling major national staffing contracts in record-breaking turnaround times, providing clients with the management information they need to see, achieving less unscheduled delays and more reliable journeys, better worker safety, rail maintenance through enhanced predictive and scheduled maintenance and real-time maintenance interventions as well as increased network availability.

Ports as intelligent logistics hubs may involve systems that provide unmanned, autonomous transfer of equipment, baggage, people, information or resources from point-to-point with minimal human intervention. Such hubs can target significantly improved operational efficiency, improved terminal operations, less delays, reduced energy consumption, better customer experience, streamlined supply chain, enhanced port traffic with less truck and vessel congestion.

Efficient air transport involves edge technology for significantly improved operational efficiency and proactive disruption management, less missed flight connections, less passenger wait times, reduced lost baggage, less passenger complaints, more stable ticket prices, and better airport business and retailing.

Integrated urban mobility involves a planning concept supporting the integration and balanced development of all modes for reduced traffic congestion and delays in public transport, faster freight distribution in city centres with reduced delivery mileage, less congestion, better customer satisfaction, better asset management, and improved situational awareness.

Shared Logistics for e-commerce reflects a desire for a modal shift, or a change between modes, and usually encompasses an increase in the proportion of freight trips made using sustainable modes targeting significantly improved operational efficiency, optimised capacity utilisation, increased consumer satisfaction, less waiting time for deliveries, and enhanced retailing and e-commerce.

[93] https://transformingtransport.eu/transport-domains

In the project's initial phase, concrete areas of assessment were discussed with regard to the impact of the different technologies to be tested.[94] Through the research of scientific literature, previous experience in other projects of scenario testing in the transportation sector and discussion with project partners and experts in the field a set of six general categories for the horizontal assessment were established: Operational efficiency, asset management, environmental quality, energy consumption, safety and economic considerations. Privacy and data protection do not play a key role within the scope of the impact assessment.

The TT consortium acknowledges that in modern society more and more data is used for complex services.[95] It is stated that the processing of information is necessary to enable the big data analytics in services. This leads to additional responsibilities in modern business models. Companies and in particular business models increasingly depend upon confidential data such as intellectual property, market intelligence and personal data. The challenge is to maintain the privacy and confidentiality of this data as well as meeting the requirements of a growing list of related compliance obligations. Therefore, TT created a data governance guideline. The guideline is based on other European guidelines and laws regarding data privacy.

Best practices are adopted with respect to the clarification of requirements and goals, making sure the right data asset is used, the clarification of the license for usage and the validation of the availability date. The guideline does not only provide general recommendations related to data assets but goes into detail with respect to the anonymisation of personal data, the aggregation of data, the homogenisation of structured data, licensing, availability, costs as well as legal and quality issues. The general recommendations deal with integrity, transparency, accountability, auditability, checks and balances, stewardship, standardisation and change management.

3.2.4. Other projects

There are several other projects that focus on transport and smart city scenarios. Among them are, for instance, QROWD, CLASS, TYPHON and Track & Know.

The QROWD[96] project offers innovative solutions to improve mobility, reduce traffic congestion and make navigation safer and more efficient. To achieve that, QROWD integrates geographic, transport, meteorological, cross-domain and news data with the goal of maximising the value of big data in planning and managing urban traffic and mobility. QROWD's platform and technologies feature a dedicated privacy model. From a technical point of view, all the collected data is linked with a random unique identifier that does not allow to link the data with the user who generated it. Additionally, all the

[94] D2.2 Analysis of Pilot Requirements for Big Data Use, https://transformingtransport.eu/sites/default/files/2017-08/D2.2%20%E2%80%93%20Analysis%20of%20Pilot%20Requirements%20for%20Big%20Data%20Use.pdf
[95] D1.3 Data Management Approach, https://transformingtransport.eu/sites/default/files/2017-08/D1.3_Data_Management_Approach_V2.0.pdf
[96] http://qrowd-project.eu/

raw data are stored for a limited amount of time.[97] Furthermore, QROWD has an ethical advisory board that aims to ensure that the project's data collection and processing guidelines were in line with TomTom's and Comune di Trento´s privacy and data protection directives.[98] The consortium stresses that data is collected in compliance with the EU privacy laws.[99]

The CLASS project develops a software architecture framework to efficiently distribute big data workloads along the compute continuum. This is done in a completely transparent way, while providing sound real-time guarantees on end-to-end data analytics responses. The capabilities of the CLASS project are demonstrated in a real smart city use case with a heavy sensor infrastructure and three connected vehicles equipped with heterogeneous sensors/actuators.

The TYPHON project provides an industry-validated methodology and integrated technical offering for data persistence architectures that meet the growing scalability and heterogeneity requirements of the European industry. The project deals with use cases in domains such as automotive and motorway operation.

The Track & Know project seeks to increase the efficiency of big data, for instance, in the transport, mobility and motor insurance sectors by creating a scalable, fault-tolerant platform for big data. The project suggests a multi-disciplinary approach considering the needs of researchers, customers and the producers and providers of services. Track & Know focusses its development of software and efficient, interoperable and scalable toolboxes on automotive transportation and urban mobility in general.

3.2.5. Recommendations

We observed that projects are generally quite aware of the impacts that their big data solutions have on society. The extent of attention projects pay to societal and ethical issues, however, varies greatly. This was particularly obvious for the projects focusing on the transportation and smart city context, but it may also be true for other contexts including healthcare and web browsing. Projects usually assess the impacts of their activities but usually not in a highly systematic way or following a rigorous methodology. Some establish ethics advisory boards (EAB) to make sure that ethical and societal aspects are considered. Most of the projects with an EAB involve external advisors in order to benefit from third-party perspectives. Failure to asses and adequately address impacts can result not only in considerable damage to a project's (or an organisation's) reputation and finances but also in harm or disadvantages for individuals. To address this problem, impact assessments have to follow systematic methodologies. The Art. 29 Working Party's Guidelines on DPIA as well as ISO/IEC 29134:2017 may be particularly useful to implement systematic impact assessments as they are very well structures and comprehensive. They may be complimented by the UN's Data Innovation Risk Assessment Tool, DataEthics, the Data Ethics Canvas, SoDIS or the e-SIDES Applied Ethics Toolkit since these tools help making sure that the right questions are asked during the assessment. The adoption of best practices is very common among projects as this allows benefiting from the lessons learned by others. We assume

[97] http://qrowd-project.eu/wp-content/uploads/2018/07/8_Platform-and-Technologies.pdf
[98] http://qrowd-project.eu/qrowd-ethical-advisory-board/
[99] Business case requirements and design, http://qrowd-project.eu/wp-content/uploads/2018/01/Business-case-requirements-and-design.pdf

that documented best practices regarding the use of systematic, methodologically rigorous and tool-supported impact assessments would be extremely useful to increase awareness of societal and ethical issues faced in the context of big data innovation and to make sure such issues are more often and more effectively addressed than they are today.

3.3. Web browsing & third-party tracking

The big data context of web browsing and third-party tracking permeates a wide variety of uses of the Internet. As many websites and applications are funded through the sale of advertisements, technologies have grown to make these advertisements more effective. A common way in which this is currently performed is through the use third-party tracking, where third-party cookies are used to monitor user behaviour across sites, usually with the intention to present targeted or more relevant ads to the user.[100] Tracking user behaviour online can be traced back to the 1990s; however, the industry has grown to include large companies such as Google in 2009.[101] Many sites also collect large amounts of data on their own users themselves; Facebook, for instance, tracks their users' cursor movements on their website.[102] Further, third party tracking has grown to collect user data from a large number of mobile phone applications;[103] thus, the practice is a bit broader than only web browsing, and mobile phone tracking will be discussed in this section as well.

In this ecosystem, publishers own or operate websites or applications and are paid to places ads on them. The ad-network or broker is the party that connects advertisers, publishers, and users. They take advertisements and attributes for user targeting (e.g. demographics and keywords of interests) collected from advertisers, and personal data collected from users through publishers in order to place advertisements on the latter's websites or mobile applications.[104] The ad-network then conducts auctions to determine which ads are delivered to the user. The ad-network also collects the view and click reports from user interaction for billing advertisers and for sharing part of the revenue with publishers.

[100] V. Toubiana et al., Adnostic: Privacy Preserving Targeted Advertising, Adnostic Whitepaper (Stanford, NYU), https://crypto.stanford.edu/adnostic/adnostic-ndss.pdf.
[101] Google announced in March 2009 that its AdSense service would begin to present ads based on a user's behaviour while browsing. Ibid. See also, Google adsense. https://adsense.blogspot.com/2009/03/driving-monetization-with- ads-that.html
[102] J. Kanter, "Facebook is tracking you in ways you never knew — here's the crazy amount of data it sucks up", Business Insider, https://www.businessinsider.com/facebook-reveals-all-the-way-it-tracks-user-behaviour-2018-6/?international=true&r=US.
[103] R. Binns et al., Third Party Tracking in the Mobile Ecosystem, https://arxiv.org/pdf/1804.03603.pdf.
[104] In some cases, the publisher and ad-network can be the same company. For example, Facebook operates the Audience Network and Google operates AdWords.

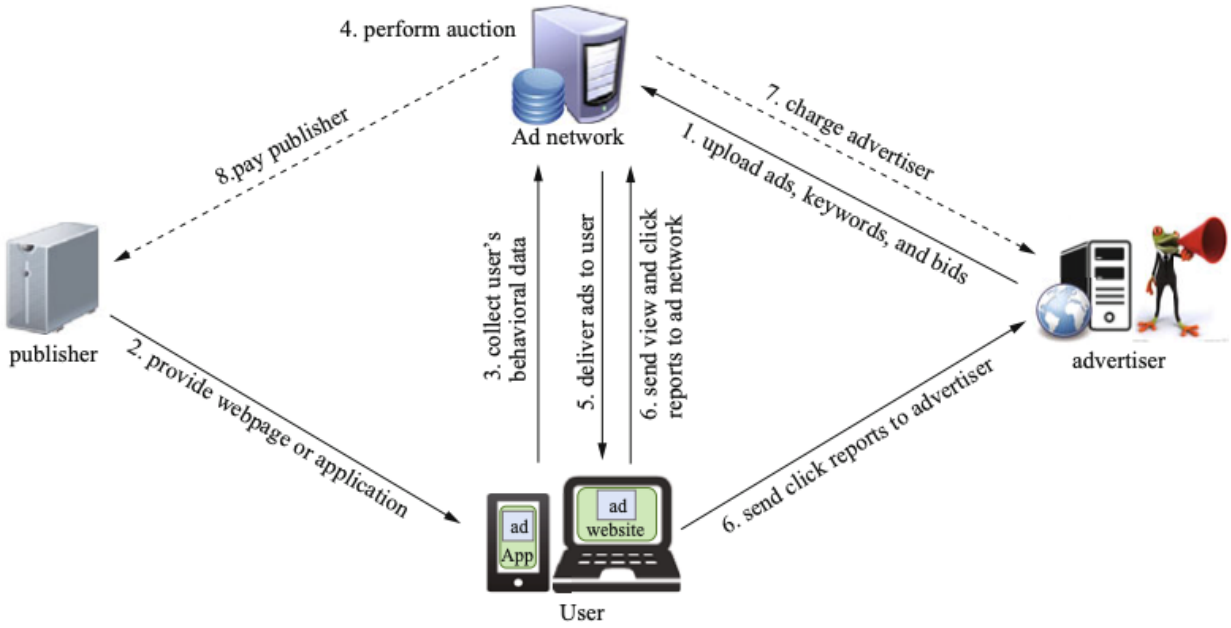

Figure 2: Targeted Advertising Structure[105]

Whether a company is operating as a publisher (website or mobile application), as a third-party ad network/data broker, or as an advertiser, the collection, storage, use, marketing, sale or purchase of massive amounts of user data creates ethical, legal, societal, and economic issues. These issues warrant the use of one of the aforementioned impact assessment tools, which can help companies anticipate problems before they arise and to address unforeseen complications in an efficient manner. While there already exist a number of privacy-preserving technologies to aid in addressing some of these issues, further technologies are in development to better ensure security and that the end users' rights are being upheld, several of which will be examined below. The introduction of these privacy-preserving technologies into a big data solution would also warrant the use of assessment tools to evaluate their impact. However, as many of these privacy-preserving technologies are still in early stages of development, they have not been thoroughly assessed using such tools, and thus such an assessment would fall on the company looking to introduce the PPT as many aspects will be context dependent. Nonetheless, this section will point out several ethical, legal, economic, and societal implications of the use of the technologies.

Some key observations we aggregated about this context are the following:

- Little to no big data stakeholders have implemented impact assessment tools.
- Users may have to take a stronger stand in order to protect their privacy.
- Ad-networks operate in the shadows and have not had much enforcement from DPAs.

[105] J. Jiang et al., "Towards privacy-preserving user targeting", Journal of Communications and Information Networks, Vol.1, No.4, Dec. 2016.

3.3.1. Ad-networks

It has been noted that it is extremely difficult to optimise the three design goals of 1) ad relevance, 2) privacy, and 3) efficiency in a single personalised advertisement system.[106] Because of this, many of the technologies below sacrifice ad relevance or efficiency (or both) at the expense of privacy.

As most ad-networks collect and analyse users' personal data on their servers, one common technical approach to preserve the privacy of users is by limiting the data collected. These client-based solutions generally keep the user data on their device and then perform targeting locally, but generally require the participation of the ad-network.

One form of this, called hybrid personalisation mechanisms, downloads a number of ads from ad-networks in a variety of areas, and then the most appropriate ad is then displayed to users based on their data stored locally.[107] Examples of this approach include Adnostic,[108] Privad,[109] and MobiAd.[110]

Adnostic uses a Firefox browser extension, one module builds local user profiles locally, and the other downloads a number of ads and then selects which of them to display to the user. User click data is recorded and sent to the ad-networks in a way that preserves the privacy of users; this requires the use of zero-knowledge proofs, homomorphic encryption and the use of a trusted third party. However, one major impediment to the use of this technology is that it will only function with the cooperation of ad-networks, and changes would be needed in the way they serve ads.

Privad works somewhat similarly. It downloads all potential ads from the broker, then chooses relevant ads based on the client. A 'dealer' functions as an anonymous proxy between the client and broker, which prevents the identification of the client or user. Connections between the client and dealer are encrypted, and thus the dealer cannot obtain data about which ads were downloaded. A 'monitor' is used to ensure that the client is not transmitting data to the broker covertly. Privad also incorporates a manual selection mechanism, where users can subscribe to categories of advertisements they might find interesting. Privad does not 'trust' ad-networks, and anonymises all data sent from the client. This function impacts performance, and the detection of fraudulent ad clicks would require additional resources as well, and thus advertisers are unlikely to prefer such an option.

---

[106] M. Hardt and S. Nath. "Privacy-aware personalization for mobile advertising". In: *Proceedings of the 2012 ACM conference on Computer and communications security (CCS 2012)*. ACM. Raleigh, NC, USA, 2012, pp. 662–673.
[107] Y. Liu, 'Privacy-Preserving Targeted Advertising for Mobile Devices' (2017).
[108] V. Toubiana et al. "Adnostic: Privacy preserving targeted advertising." In: *Proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS 2010)*. Retrieved April 6, 2016 from https://www.isoc.org/isoc/conferences/ndss/10/pdf/05.pdf. San Diego, CA, USA, 2010.
[109] S. Guha et al. "Privad: Practical Privacy in Online Advertising." In: *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation (NSDI 2011)*. Boston, MA, USA: ACM, New York, 2011, pp. 169–182.
[110] H. Haddadi et al. "MobiAd: Private and Scalable Mobile Advertising". In: *Proceedings of the 5th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2010)*. Chicago, Illinois, USA: ACM, New York, 2010, pp. 33–38.

Another solution makes use of special hardware. OblivAd proposes an innovative approach that requires the use of specialised hardware by the ad-network, which operates as a black box.[111] The black box distributes ads to users, and receives click and view data for each ad. It does not identify individual users, and only releases aggregated records to the ad-network. However, OblivAd is only compatible with certain ad-networks, and those that are not would require (likely expensive) changes to infrastructure to be able to implement it.

There are also a number of manual intervention solutions, such as RePriv (browser or mobile OS-based),[112] a solution by Hardt and Nath (mobile),[113] and Lockr (social networks).[114] These generally give users fine-grained control over the data they share with ad-networks, which makes the sharing of data more accurate and flexible. Some of them require the consent of the user every time personal data is requested, while others allow for the use of configuration files to be stored and used for different service suppliers. While such solutions would likely easily be compliant with legal requirements such as the GDPR, they may be cumbersome for users to employ, as they may not want so many prompts for their consent. The different technologies use a variety of methods to mitigate this concern, so that prompts are more or less answered automatically.

PAPAYA (Platform for privacy-preserving data analytics)[115] is a Horizon 2020 project that aims to address privacy concerns when data analytics tasks are performed by untrusted third-party data processors. PAPAYA is designing and developing dedicated privacy preserving data analytics modules that will enable data owners to extract valuable information from encrypted data, while being cost-effective and accurate. Ultimately, this technology will be able to be used in a variety of contexts, such as web and mobile data and e-Health.

3.3.2. Users

Other technologies, called context obfuscation solutions, do not share any personal data with online service providers. For example, with NOYB[116] (none of your business), researchers encrypted user data, which is then encoded to appear as legitimate data. The online service can use that data to serve ads, but cannot discover the user through the use of this synthetic data. As users may be encoded so as to appear as belonging to a different gender and age range, they may be served less relevant ads,

[111] M. Backes et al. “Obliviad: Provably secure and practical online behavioral advertising”. In: *IEEE Symposium on Security and Privacy*. IEEE. 2012, pp. 257–271.
[112] M. Fredrikson and B. Livshits. “Repriv: Re-imagining content personalization and in-browser privacy”. In: *Proceedings of IEEE Symposium on Security and Privacy (SP 2011)*. 2011, pp. 131–146.
[113] M. Hardt and S. Nath. “Privacy-aware personalization for mobile advertising”. In: *Proceedings of the 2012 ACM conference on Computer and communications security (CCS 2012)*. ACM. Raleigh, NC, USA, 2012, pp. 662–673.
[114] A. Tootoonchian et al. “Lockr: better privacy for social networks”. In: *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*. ACM. 2009, pp. 169–180. For more information see Section 3.3.3 on Publishers.
[115] PAPAYA, https://www.papaya-project.eu/content/papaya-announcement-letter.
[116] S. Guha et al. “NOYB: Privacy in online social networks”. In: *Proceedings of the 1st Workshop on Online Social Networks*. ACM. 2008, pp. 49–54.

potentially resulting in losses to consumer benefit and advertisement revenue. This technology is implemented as a Firefox extension and requires no cooperation on the part of the ad-network. Other obfuscation solutions submit synthetic data along with real data. Examples include TrackMeNot[117] and ProfileGuard,[118] the former works as a Firefox extension and submits false clicks and views, and the latter makes it appear as though apps are installed on a mobile device that in fact are not. Similarly, consumers would lose the potential benefit of personalised services and ads, while in some cases the vast amount of synthetic data generated may degrade the performance of networks.

Other tools block trackers and sometimes ads themselves. The browser extensions Privacy Badger (from the Electronic Frontier Foundation)[119] and the open-source project uBlock Origin[120] both block common trackers, and the latter filters out advertisements as well. Still others address 'active' content such as embedded Flash or JavaScript (among others), which can be used for web tracking or even malicious purposes; these tools include uMatrix[121] and NoScript.[122]

There are also sandbox solutions for ads in order to protect personal data; these function similarly to sandboxes that work in other technological environments that allow users to run programs and code from unverified suppliers safely. AdJail[123] and AdSentry[124] are two options that sandbox advertisements for browsers. Other sandboxing mechanisms are available for mobile Android devices, such as Mobile-sandbox[125] and DroidBox,[126] the latter of which also uses TaintDroid[127] to monitor untrusted app and analyse how these apps are using the user's personal data.

3.3.3. Publishers

Most of the privacy-preserving technologies in development generally focus on user-client and ad-network solutions (or a combination of both), with the website or mobile app publisher as an

[117] D.C. Howe and H. Nissenbaum. "TrackMeNot: Resisting surveillance in web search". In: *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* 23 (2009), pp. 417–436.
[118] I. Ullah et al. "ProfileGuard: Privacy Preserving Obfuscation for Mobile User Profiles". In: *Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES 2014)*. ACM. 2014, pp. 83–92.
[119] Electronic Frontier Foundation, "Privacy Badger", https://www.eff.org/privacybadger.
[120] uBlock Origin, https://github.com/gorhill/uBlock.
[121] uMatrix, https://github.com/gorhill/uMatrix.
[122] NoScript, https://noscript.net//
[123] M. Ter Louw, et al. "AdJail: Practical Enforcement of Confidentiality and Integrity Policies on Web Advertisements". In: USENIX Security Symposium (2010), pp. 371–388.
[124] X. Dong, et al. "AdSentry: Comprehensive and flexible confinement of JavaScript-based advertisements". In: Proceedings of the 27th Annual Computer Security Applications Conference (2011), pp. 297–306.
[125] M. Spreitzenbarth, et al. "Mobile-sandbox: Having a deeper look into Android applications". In: Proceedings of the 28th Annual ACM Symposium on Applied Computing (2013), pp. 1808–1815.
[126] A. Desnos and P. Lantz. "Droidbox: An Android application sandbox for dynamic analysis", http://www.honeynet.org/gsoc2011/slot5.
[127] W. Enck, et al. "TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones". In: ACM Transactions on Computer Systems (TOCS) 32.2 (2014), p. 5.

intermediary. However, there is most often no connection between users and ad-networks, and publishers thus play an important role and often have related legal obligations.

One technological approach mentioned above can be applied by publishers. Lockr[128] gives users control over the information they share with social networks, including whether those networks can store it and which third parties can access it. Other features include mechanisms to prevent data from being reused for unintended purposes, and uses encryption to enable users to find and add friends to their networks without revealing their relationships to one another.

As for the legal obligations, it is considered that the legal basis for organisations processing personal data related to tracking cookies (and other similar technologies) under the GDPR must be the *consent* of the user or data subject.[129] "Legitimate business interests" as a legal basis is unlikely to be able to be used due to balancing of interests that takes place (i.e. the right to privacy and data protection outweigh the right to do business).[130] Obtaining consent coincides with other requirements under the e-Privacy Directive and the GDPR.

When cookies are used, as they often where targeted advertising is used, the website publisher must provide users "with clear and comprehensive information, in accordance with Directive 95/46/EC [replaced by the GDPR], *inter alia*, about the purposes of the processing" under the e-Privacy Directive.[131] Users must consent to this notice for the cookies to be used, and they must be offered the right to refuse.[132] Obtaining consent while providing the proper information can be difficult: there is a tension between providing clear and concise information while still explaining how one's data is actually collected and used.

Even though the cookie notification requirements were in place for nearly a decade before the GDPR came into force, many websites are still uncompliant. However, the higher potential fines under the GDPR likely influenced more websites to add notifications.[133] While an increasing number of websites allow users to accept or block certain types of cookies,[134] tracking purposes can get lost among the large

---

[128] A. Tootoonchian et al. "Lockr: better privacy for social networks". In: *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*. ACM. 2009, pp. 169–180.

[129] F.J.Z. Borgesius and J. Poort, "Online Price Discrimination and EU Data Privacy Law", *Journal of Consumer Policy*, Vol. 40, 347, 360-61 (2017); Frederik J. Zuiderveen Borgesius, 'Personal data processing for behavioural targeting: which legal basis?' *International Data Privacy Law*, Vol. 5, No. 3, 163 (2015).

[130] Ibid.

[131] Directive 2002/58/EC, of the European Parliament and of the council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), 2002 O.J. (L 201) 37 [hereinafter ePrivacy Directive]. The GDPR requires that this information be in a "concise, transparent, intelligible and easily accessible form, using clear and plain language". GDPR, Art. 12(1).

[132] ePrivacy Directive, Art. 5(3).

[133] The use of cookie consent notifications increased from 46.1% in January 2018 to 62.1% in May 2018 on websites in the EU. Degeling et al., "We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy", p. 6, arXiv:1808.05096v2.

[134] In a 2015 study by the Article 29 Working Party, only 16% of websites surveyed offered such options. Article 29 Data Protection Working Party, *Cookie Sweep Combined Analysis – Report*, at 20, 14/EN, WP 229 (3 February 2015).

number of other purposes that cookies are used for. Moreover, depending on how the cookie notification banner functions, it can lead to questions of whether there is informed consent if there is an easy 'accept all' button while opting out is made difficult.[135]

Cumbersome cookie notification banners have led to users becoming frustrated by them, and the upcoming ePrivacy Regulation aims to help users opt out of certain cookies through the use of settings in their browsers, and by introducing stricter rules on tracking.[136]

Some of the aforementioned technologies in development may be considered 'extreme' in that they completely undermine the current targeted advertising model. This could be seen as a response to the advertising industry's lack of response to users' desires to opt-out of tracking. For instance, most browsers have supported a technology called 'Do Not Track' for years; the browser of a user who has it enabled requests websites to not track them.[137] However, the technology requires the cooperation of the website or publisher, and most sites do not honour the request.[138] In general, the websites/publishers and ad-networks/brokers did not want to give up the ad revenue and detailed personal data upon which the revenue was based.

Websites and mobile app developers are not just intermediaries in this scheme, and the decisions they make can affect the privacy of its users. After conducting impact assessments, these publishers may decide that it is best to refrain from partnering with the ad-networks that use more invasive tracking, thus protecting more privacy of its users. Some companies have done exactly this; they have implemented ethical committees that also do legal research, which investigate whether to partner with certain organisations based on their data use and then to monitor whether their partners are using the shared data as stated. Because this may affect them economically, they may have to turn to alternative business models, perhaps charging users one-time, subscription, or micropayment fees.[139]

3.3.4. Recommendations

The privacy-preserving technologies discussed above are in different stages of development. Many of them are theoretical and have not been explored outside of their introduction in an academic article. Others are being actively developed and used. The divide appears to occur between solutions that require the participation of ad-networks and those that can be implemented solely by users, with the

[135] One tactic used is that the user must individually click each ad-network tracker in order to avoid tracking.
[136] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) (10 January 2017).
[137] K. Hill. "'Do Not Track,' the Privacy Tool Used by Millions of People, Doesn't Do Anything", *Gizmodo* (15 October 2018), https://gizmodo.com/do-not-track-the-privacy-tool-used-by-millions-of-peop-1828868324.
[138] Ibid.
[139] A number of micropayment options exist, including one from Paypal, and development using cryptocurrencies has been proposed. *See* S. Jain and A. Narayanan, "Monetization on the Modern Web: Automated Micropayments From Bitcoin-Enabled Browsers" (2016).

user-oriented solutions being the most actively developed, with some tools being used by millions of users.[140]

As people become increasingly aware of these tracking technologies and start to use technologies to counter them, it might be in the best interest of the ad-networks and publishers to work towards cooperative solutions—even if there are costs in implementation and efficiency—that provide users control over their privacy while still allowing them to provide personalised features or advertisements. Not only would such an approach likely be economically prudent in the long run, the increased transparency and informed consent would also aid in complying with ethical norms and the legal requirements noted above.

We observed in this context as to which impact assessment tools shall be best implemented depends on the stakeholder as well as a variety of factors, such as the fact that some tools might not be compatible with how their systems are set up.

Nevertheless during setting up such systems they could benefit from applying the Art. 29 Working Party Guidelines on DPIA, the Data ethics impact assessment and guidelines, the ISO/IEC 29134:2017 and potentially e-SIDES Applied Ethics Toolkit. The two latter tools seem especially useful, because web-browsing is a highly economically-oriented context from multiple perspectives (retailer, purchaser/user, advertiser, etc.), and the last two tools include a step-by-step practical guide for stakeholders and how to assess the implications of implementing privacy-preserving technologies and big data solutions in general.

[140] uBlock Origin is used by more than 10 million users of Chrome (https://chrome.google.com/webstore/detail/ublock-origin/cjpalhdlnbpafiamejdnhcphjbkeiagm?hl=en), and by more than 4.5 million users of Firefox (https://addons.mozilla.org/en-US/firefox/addon/ublock-origin/).

4. Conclusion

This deliverable posed the research question: How are security and privacy features in big data-driven innovation projects—which we call privacy-preserving technologies—suitable to deal with the ethical, legal and social values that are under pressure in specific big data application contexts? The answer to this research question also includes answers on whether the embedded features add new issues and put additional values at risk. For this purpose, we identified ethical, legal and societal impact assessment tools. We also identified three big data contexts: healthcare, transportation and smart cities, and web browsing and third-party tracking, in order to narrow the scope and increase the depth of our assessment. On the basis of answers upon our survey and desk research, we concluded the following:

- Privacy-preserving technologies are diverse, yet a common denominator for them is that they often result in trade-offs between privacy, confidentiality interests and business interests. Through compromising towards privacy, the granularity of data for business purposes becomes difficult to exploit in all three big data contexts.

- Insights on the ethical, legal and societal impacts assessment tools regarding the BodyPass, MyHealthMy Data, CLARUS and BigMedilytics projects show that the healthcare sector is scattered. Despite the fact that the strictest data protection rules apply for healthcare data, diverse privacy-preserving technologies and types are implemented for epidemiological, diagnostic and other reasons. Each of the four projects could benefit from relying upon one or more of the impact assessment tools we identified before.

- Within the context of transportation and regarding the SPECIAL, AEGIS, and Transforming Transport projects, companies and business models increasingly rely upon confidential data such as intellectual property, market intelligence and personal data. In order to overcome the challenge, we found that maintaining the privacy and confidentiality of data and meeting compliance requirement, such as the data governance guideline created by the Transforming Transport project, for instance. In terms of impact assessments, best practices are adopted with respect to the clarification of requirements and goals, ensuring the right data asset is used, the clarification of the license for usage and the validation of the availability date. The guideline does not only provide general recommendations.

- A wide variety of different privacy-preserving technologies are in development in the web browsing and third-party tracking context. Some of the technologies can be employed unilaterally by one of the relevant parties (ad-networks, users, and to a lesser extent publishers), while other solutions require cooperation between these parties. Generally, those technologies that require the participation of ad-networks have not seen much utilisation perhaps due to costs and lost revenue, while unilateral technologies employed by users to protect their privacy are actively developed and used. As this context develops, cooperative solutions that protect the privacy of individuals, while still being able to provide personalised features and advertisements offer the most beneficial way forward.

As a general conclusion, it can be pointed out that finding an adequate balance between the protection of privacy, confidentiality and the utility of big data is necessary but complicated. This was an issue identified in all project that were investigated. The ICT-14, ICT-15 and ICT-16 projects we assessed had different approaches with respect to how to protect privacy by developing privacy-preserving technologies and accompanying mechanisms. The evaluation of the projects has shown that, depending on their context, purpose and whether they relied themselves upon one or multiple privacy-preserving technology within the big data value chain[141] the assessment of impact would require different and often combined impact assessment tools. The SPECIAL project, for instance, developed an environment for managing personal data usage policies, in which the assessment of the impact of the involved technologies required a combination of mechanisms. For instance, impact assessments needed to ensure that companies comply with legislation and usage policies defined by data subjects. This increased the extent to which SPECIAL's extensible environment lived up to the promises of ensuring more visibility on the processing of data of data subjects.

The three big data contexts demonstrated that big data-driven innovations bring along fast change in professional, public, private and commercial relationships of citizens and this requires the continuous assessment of impacts not only with respect to the privacy of individuals but also to the accountability of data sharing processes. The seven impact assessment tools could therefore be regarded as useful in all three contexts but to differing degrees. Furthermore, the fast pace of innovative developments and the variety of ways in which privacy-preserving technologies are used and combined also underline the aspect of time during the assessment of impacts. From all three contexts it became apparent that privacy-preserving technologies and data sharing methods serve the optimisation of processes that are increasingly geared toward improving predictions on the future behaviour of citizens. Hence, beyond the continuity in assessing the present impact of technologies, impact assessments should also embrace the assessment of potential impacts in the future.

[141] This chain includes different stakeholders that are involved in the exchange and analysis of data.

Appendix: related projects

This section describes related research projects and outlines aspects that are particularly relevant from the e-SIDES perspective. It is an update and extension of an earlier analysis provided in section 5 of the e-SIDES deliverable D3.1, where particular attention was paid to links between project and different classes of privacy-preserving technologies.

The information provided is based on analyses of the project websites and public reports as well as discussions with project representatives. The focus of this section is limited to projects funded by the EU. Since its launch, the e-SIDES project has taken steps to establish and maintain co-operations with related projects.

In principle, there are two types of projects that are relevant for e-SIDES: projects that develop big data solutions and projects that support the development and use of big data solutions in one way or the other. With respect to projects of the former type, we were interested in the role that ethical and societal aspects have played during the design and evaluation of the solutions. What has been developed so far? How does the evaluation look like? With respect to projects of the later type, we analysed how and under what circumstanced they have supported or plan to support the development of big data solutions and what role ethical and societal aspects have played for them. While some project help addressing ethical and societal aspects that may be faced by developers of big data solutions, others provide more general help only if the developers have taken measures to ensure their results are ethical and socially compatible. Forms of support may be consulting services, guidelines, solutions for certain technical problems or financial benefits.

ICT-18-2016 RIA

The focus of this section is on the Research and Innovation Actions (RIAs) SPECIAL, SODA and MHMD. Just like the Coordination and Support Action (CSA) e-SIDES, these projects are funded under ICT-18-2016 (Big data PPP: Privacy-preserving big data technologies). The three RIAs are tasked to advance technologies for data access, processing and analysis to better protect personal data in line with existing and future EU rules. All three project are in their final year.

SPECIAL

The SPECIAL[142] (Scalable Policy-aware Linked Data Architecture for Privacy, Transparency and Compliance) project aims to address the contradiction between big data innovation and data protection by proposing a technical solution that makes both of these goals more realistic. The project was discussed already at length as part of the analysis of related projects provided in D3.1.

The technical solution, the SPECIAL platform, is an extensible environment for managing personal data usage policies, ensuring compliance with such policies and tracking personal data usage along with the context it is being used in. In April 2019, a release of the platform was made available that provides working implementations of many of the ideas presented recent deliverables (i.e., D2.5, D2.7, D2.8).

[142] https://www.specialprivacy.eu/

Moreover, the new release brings updates with respect to ex ante compliance checking, consent backend changes, the personal data inventory, and compression and encryption.

D1.7, which was published after the previous analysis of SPECIAL, includes details on the policy and transparency considerations of the use case, and the compliance requirements against policies. D1.7 is particularly relevant from the e-SIDES perspective.

The privacy-preserving technologies used in the SPECIAL context as well as the assessments tools considered for evaluating the impact of the project's technologies were discussed in more detail in section 3.2.1.

The 3-year project started in January 2017 and will end in December 2019.

SODA

The **SODA**[143] (Scalable Oblivious Data Analytics) project has two main aims. The first aim is to enable practical privacy-preserving analytics of information from multiple data assets using MPC techniques. The second aim is to combine MPC with a multidisciplinary approach towards privacy. The project was also discussed at length in D3.1.

SODA has made contributions to different MPC frameworks such as FRESCO, MPyC and PySNARK. FRESCO is a Java framework for efficient secure computation aiming at making the development of prototype applications based on secure computation easy. SODA has made substantial improvements and expansions to the FRESCO framework enhancing stability, usability and security. MPyC is a Python framework for secure multiparty computation. MPyC targets usability and ease of use balanced with efficiency. These properties make it well-suited for education and rapid prototyping of MPC applications. SODA implemented various machine learning algorithms in FRESCO and MPyC. PySNARK is a Python-based system for zk-SNARK (zero-knowledge succinct non-interactive argument of knowledge) based verifiable computations and smart contracts. PySNARK makes it possible to program verifiable computations in Python. This solves the problem that it is quite hard to specify verifiable computations. PySNARK automatically creates proofs.

D2.1 and D3.1, two deliverables that are particularly relevant for e-SIDES, were not yet publicly available at the time the previous analysis of the project was conducted. D2.1 contains a state-of-the-art analysis of MPC-based privacy-preserving data mining protocols used in the big data context. D3.1 provides a thorough legal analysis of the current privacy law in the EU, with emphasis on the GDPR.

The 3-year project started in January 2017 and will end in December 2019.

MHMD

The **MHMD**[144] (My Health - My Data) project aims to build and test new models of privacy and data protection that meet the requirements of the biomedical sector, in which issues of data subjects' privacy

[143] https://www.soda-project.eu/
[144] http://www.myhealthmydata.eu/

and data security represent a crucial challenge. Just as the other two ICT-18 projects, MHMD was discussed at length in D3.1.

With respect to deliverables that are particularly relevant from the e-SIDES perspective, MHMD published D3.2 after the previous analysis of the project. The deliverable describes the evaluation framework for the management of patient consent and its implementation as a smart contract. The framework serves as basis for getting approval from the project's partner hospitals' ethics committees. Therefore, it does not only document the dynamic consent principle but also details the ways in which consent is implemented on the blockchain in code, shared with external parties, and automatically self-enacted and possibly modified by the patient.

The privacy-preserving technologies used in the MHMD context as well as the assessments tools considered for evaluating the impact of the project's technologies are discussed in more detail in section 3.1.2.

The 3-year project started already in November 2016 and will end in October 2019.

ICT-14-2016-2017 IAs

In this section, we focus on the Innovation Actions (IAs) funded under ICT-14-2016-2017 (Big data PPP: Cross-sectorial and cross-lingual data integration and experimentation). The IAs are tasked to address data challenges in cross-domain setups. Particular attention is paid to the projects that are in their final year: SLIPO, EW-Shopp, QROWD, euBusinessGraph, FashionBrain, BigDataOcean and Data Pitch. Data Pitch serves as an incubator for data-driven start-ups, while the other projects develop big data solutions themselves. These projects have already been part of the previous analysis of related projects. Moreover, we analysed the projects that are in their second year: TheyBuyForYou, Lynx, ICARUS, BodyPass, FANDANGO, Cross-CPP and EDI. All but EDI, which serves as an incubator, develop big data solutions.

SLIPO

The **SLIPO**[145] (Scalable Linking and Integration of scalable Big POI data) project aims to deliver technologies to address the data integration challenges of POI data in terms of coverage, timeliness, accuracy and richness. SLIPO focuses on the transfer of the research output of previous projects to POI data and the introduction of validated and cost-effective innovations across their value chain. SLIPO extends and integrates open source software for semantic integration of geospatial data, aiming to develop effective, efficient and scalable tools for POI integration and enrichment.[146] As the project does not use personal data, privacy is not a key issue for SLIPO. However, trustworthiness of data is a central aspect, according to a project representative.[147]

The 3-year project started in January 2017 and will end in December 2019.

[145] http://www.slipo.eu/
[146] http://www.slipo.eu/?page_id=29
[147] A project representative was interviewed during the European Big Data Value Forum of 2017.

AEGIS

The **AEGIS**[148] (Advanced Big Data Value Chain for Public Safety and Personal Security) project aims at creating an interlinked "Public Safety and Personal Security" (PSPS) data value chain, and at delivering a novel platform for big data curation, integration, analysis and intelligence sharing. AEGIS will help EU companies to adopt a more data-driven mentality, extending and/or modifying their individual data solutions and offering more advanced data services (e.g., data cleansing, data integration, semantic data linking) while at the same time attaching value to their datasets and introducing novel business models for the data sharing economy.

Through the AEGIS platform, the PSPS data value chain analysis is conducted at multiple levels: (I) data privacy enhancement, (II) data pre-processing, (III) big data analysis, and (IV) data intelligence sharing.[149] It also develops an automotive and road safety demonstrator, a smart home and assisted living demonstrator and an insurance sector demonstrator.[150]

Consideration of ethical and societal aspects is an important part of the AEGIS project. It has an own ethics advisory board (EAB) working closely with the AEGIS consortium. In D1.3, there is a chapter on ethical, privacy, data protection and IPR strategy that outlines an own privacy concept, acknowledges the European Convention of Human Rights and the Charter of Fundamental Rights of the EU, and also the GDPR and the Directive 2002/58/EC (ePrivacy Directive) are considered[151]. Other deliverables deal with anonymisation and data cleansing tools that can be controlled by the user.[152] D9.1 is dedicated to ethical issues and defines a strategy for the assessment of data protection and ethical issues, and also applies it to the three demonstrators developed in the project.[153] In D9.3, activities, assessments and recommendations of the EAB are described.[154]

In April 2019, AEGIS gave a presentation within the scope of an e-SIDES event. The project's approach towards requirements elicitation was outlined. Regarding privacy by design and by default, privacy principles are embedded into the design process of data processing systems since the very beginning. Regarding its privacy protection goal, the key role of private individual's point of view and protection goals as central element for deriving requirements, as well as for identifying risks, countermeasures and for assessment are described. AEGIS defines security protection goals (confidentiality, integrity, and

---

[148] https://www.aegis-bigdata.eu/
[149] https://www.aegis-bigdata.eu/approach/
[150] https://www.aegis-bigdata.eu/demonstrators/
[151] https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D1.3-Final-AEGIS-Methodology-v1.0.pdf
[152] D3.2 - AEGIS Components, Microservices and APIs Design v1.00, 2017-11-30, https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D3.2-AEGIS-Components-Microservices-and-APIs-Design-v1.00.pdf; D3.3 - Architecture and Revised Components, Microservices and APIs Design v2.00, 2018-07-19, https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D3.3-Architecture-and-Revised-Components-Microservices-and-APIs-Designs-v2.00_v1.1.pdf; D4.2 - AEGIS Platform - Release 2.00, 2018-07-17, https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D4.2-AEGIS-Platform-Release-2.00-v1.0.pdf
[153] https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D9.1-Other-Ethical-Issues_v1.0-public.pdf
[154] https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D9.3-GEN-Requirement-N-%C2%B0-4-_EABs-Report_v1.0_public.pdf

availability) and further specific privacy protection goals (unlinkability, transparency, intervenability). The project promotes the balance of privacy and security requirements against other protection goals.[155]

A more comprehensive assessment of the AEGIS project is provided in section 3.2.2.

The 2.5-year project started in January 2017 and will end in June 2019.

EW-Shopp

The **EW-Shopp**[156] (Supporting Event and Weather-based Data Analytics and Marketing along the Shopper Journey) project aims to deploy and host a data integration platform to ease data integration tasks in the e-commerce, retail and marketing domain. This is achieved by embedding shared data models, robust data management techniques and semantic reconciliation methods. The project integrates contextual variables (e.g., weather conditions, calendar events, holidays) into the analysis of consumer behaviour.

No new deliverables have been published since the last evaluation of the project and there is no direct relation to ethical and societal issues visible.

The 3-year project started in January 2017 and will end in December 2019.

QROWD

The **QROWD**[157] (Because Big Data Integration is Humanly Possible) project aims to offer innovative solutions to improve mobility, reduce traffic congestion and make navigation safer and more efficient. To achieve that, QROWD integrates geographic, transport, meteorological, cross-domain and news data with the goal of maximising the value of big data in planning and managing urban traffic and mobility. Among others, it develops advanced road information, intelligent urban transportation and mobility and a Virtual City Explorer.[158] Some of these are already finished.

Regarding ethical and societal aspects, anonymisation and privacy are a general topic in QROWD.[159] Its platform and technologies feature a dedicated privacy model. User data is stored and processed anonymously through a unique identifier generated randomly when the user registers on i-Log. A disambiguation table allows to retrieve the user identified, when needed. To achieve minimisation, the citizen's email (identifier) and only sensor data for mobility (GPS, gyroscope and accelerometer) are collected. Additionally, all the raw data are stored for a limited amount of time.[160] Furthermore, QROWD

[155] AEGIS presentation for e-SIDES
[156] http://www.ew-shopp.eu/
[157] http://qrowd-project.eu/
[158] http://qrowd-project.eu/wp-content/uploads/2018/07/2_Advanced_road_information_biz_case.pdf; http://qrowd-project.eu/wp-content/uploads/2018/07/3_Intelligent_urban_mobility.pdf; http://qrowd-project.eu/wp-content/uploads/2018/07/6_Human-Machine-Workflows-for-Mobility-Infrastructure.pdf
[159] http://qrowd-project.eu/wp-content/uploads/2018/12/D7.3%E2%80%93Dynamic-data-integration2c-storage-and-access-1.pdf
[160] http://qrowd-project.eu/wp-content/uploads/2018/07/8_Platform-and-Technologies.pdf

has an ethical advisory board that aims to ensure that the data collection and processing guidelines were in line with TomTom's and Comune di Trento´s privacy and data protection directives.[161]

The 3-year project started in December 2016 and will end in November 2019.

euBusinessGraph

The **euBusinessGraph**[162] (Enabling the European Business Graph for Innovative Data Products and Services) project aims to create a business graph, which is understood as a highly interconnected graph of Europe-wide company-related information both from authoritative and non-authoritative sources (including data from both the public and the private sector). By doing so, the project strives to provide a data marketplace that enables the creation of a set of data-driven products and services via a set of six business cases. It develops various services such as a corporate events data access service (CED)[163], a tender discovery service (TDS)[164] and a customer relationship service (CRM-S)[165].

There is no direct connection to ethical, societal and privacy issues in the five deliverables (however, there is reference to the GDPR) and no ethics board or similar is visible on the website and in the deliverables.

The 2.5-year project started in January 2017 and will end in June 2019.

Fashion Brain

The **FashionBrain**[166] (Understanding Europe's Fashion Data Universe) project aims at combining data from different sources to support different fashion industry players (i.e., the retailers and the customers) by predicting upcoming fashion trends from social media as well as by providing personalised recommendations and advanced fashion item search to customers.

The project puts emphasis on data ethics and privacy. In its project goals, it is stated that all participation will be anonymised and any reporting, which will be done in the aggregate, will ensure that anonymity is protected. Additionally, the anonymity of all individual users is protected by only keeping data which is essential to the project's objectives. Once collected, the data is filtered to remove spam and irrelevant content, aggregated statistics and models will be produced. All personal data that is no longer to be used for the project's purposes is immediately deleted. In the case of Instagram posts, metadata about the images and information of their classification is produced. For any data that subsequently becomes part of a sharable data set (with the permission of participants as outlined in the formal consent), the data will be anonymised to protect participants' identity before being released/transferred. Video, images, and audio data may never be released under such conditions unless anonymity can be guaranteed. In all cases, any restrictions related to existing or newly collected

[161] http://qrowd-project.eu/qrowd-ethical-advisory-board/
[162] http://eubusinessgraph.eu/
[163] http://eubusinessgraph.eu/corporate-events-data-acess-service-opencorporates-com/
[164] http://eubusinessgraph.eu/tender-discovery-service-cerved/
[165] http://eubusinessgraph.eu/crm-service-evry/
[166] https://fashionbrain-project.eu/

data will be included with the shared data for future use. Furthermore, the FashionBrain project will not use any third-party service to process the data.[167]

The project also facilitates a Consent Manager[168] and delivers an own section on data ethics and privacy on its homepage with the ethics committee, supported by the University of Sheffield Research Ethics Committee.[169]

The 3-year project started in January 2017 and will end in December 2019.

Big Data Ocean

The **BigDataOcean**[170] (Exploiting Oceans of Data for Maritime Applications) project aims to enable maritime big data scenarios through a multi-segment platform that combines data of different velocity, variety and volume under an interlinked, trusted and multilingual engine. The project capitalises on existing big data technologies but rolls out a new value chain of interrelated data streams that is expected to transform the way maritime-related industries work.

In D2.1, it was stated that of the questions related to data collection, among others, data access control, privacy and security are addressed. It is, however, difficult to evaluate to what extend ethical and societal aspects are addressed as the ethics requirements deliverable was not made public. As it is lined out in D4.2 and D4.3, the platform architecture though features an anonymiser tool for data anonymisation.[171]

The 2.5-year project started in January 2017 and will end in June 2019.

DataPitch

The **Data Pitch**[172] (Accelerating data to market) project is an open innovation programme bringing together corporate and public-sector organisations that have data with start-ups and SMEs that work with data. It is centred around a competition with several tracks which describe challenges, and a virtual accelerator programme to help start-ups and SMEs develop solutions to meet these challenges.

Regarding ethical and societal aspects, D3.1 specifies a legal and privacy toolkit. Apart from legal compliance, challenges in data sharing and reuse (types of data, access, openness vs privacy, privacy vs utility, repurposing of personal data, data sharing and reuse protocols) are described as well as various relevant laws (EU/national data protection and privacy laws, EU/national intellectual property laws,

[167] https://fashionbrain-project.eu/data-ethics-and-privacy/
[168] https://fashionbrain-project.eu/consent_manager/
[169] https://fashionbrain-project.eu/data-ethics-and-privacy/
[170] http://www.bigdataocean.eu/
[171] www.bigdataocean.eu/site/wp-content/uploads/2016/12/BigDataOcean_Platform_Architecture_Components_Design_and_APIs_-_v1.00.pdf; www.bigdataocean.eu/site/wp-content/uploads/2018/04/BigDataOcean_Platform_Architecture_Components_Design_and_APIs_-_v2.00.pdf
[172] https://datapitch.eu/

EU/national competition laws). Also, anonymisation, pseudonymisation and re-identification risks are considered.[173]

The 3-year project started in January 2017 and will end in December 2019.

Other projects

The **TheyBuyForYou**[174] (Enabling procurement data value chains for economic development, demand management, competitive markets and vendor intelligence) project aims at building a technology platform, an online toolkit and a public portal that allows suppliers, buyers, data journalists, data analysts, control authorities and regular citizens to explore and understand how public procurement decisions affect economic development, efficiencies, competitiveness and supply chains.

There is no information available on ethical and societal aspects and no deliverables have been published yet.

The 3-year project started in January 2018 and will end in December 2020.

The **Lynx**[175] (Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe) project aims at creating an ecosystem of smart cloud services to better manage compliance, based on a legal knowledge graph (LKG) which integrates and links heterogeneous compliance data sources including legislation, case law, standards and other private contracts.

Three pilots have been designed to put the Lynx LKG in action, one of them being a pilot focusing on compliance assurance services for contracts. Among others, Lynx will create an initial knowledge base around data protection by manually and automatically collecting possibly relevant documents.[176] This pilot was created because of the GDPR. According to D1.1, the Lynx project is committed to user privacy. The specific policy for the protection the users' privacy has been designed on the basis of the GDPR.[177]Also, the data management plan in D2.1 links to ethical and societal aspects and data security.[178] Methodically, the project has set up ethical guidelines.[179]

The 3-year project started in December 2017 and will end in November 2020.

The **ICARUS**[180] (Aviation-driven Data Value Chain for Diversified Global and Local Operations) project develops a first-of-a-kind transdisciplinary methodology and toolkit for integrated impact assessment so

[173] www.datapitch.eu/wp-content/uploads/2017/06/PUBLIC-LEGAL-AND-PRIVACY-TOOLKIT-VERSION-1.0-DELIVERABLE-8.1-FINAL-30-JUNE-2017.pdf
[174] http://theybuyforyou.eu/
[175] http://lynx-project.eu/
[176] http://lynx-project.eu/project/pilot1
[177] D1.1 - Functional Requirements Analysis Report, 2018-05-31, https://zenodo.org/record/1256836#.XHfAS2Mo9aQ
[178] D2.1 - Initial Data Management Plan, 2018-05-31, https://zenodo.org/record/1256834#.XHfAS2Mo9aQ
[179] D1.1 - Functional Requirements Analysis Report, 2018-05-31, https://zenodo.org/record/1256836#.XHfAS2Mo9aQ
[180] https://icarus2020.eu/

that air quality improvement, climate change mitigation and health promotion can be evaluated efficiently in key sector policies. To this aim ICARUS works on innovative tools for urban impact assessment leading to design and implement win-win strategies to improve the air quality and reduce the carbon footprint in European cities. It develops a decision support system that is aimed at assisting stakeholders in the selection, application and evaluation of the available datasets and tools for urban impact assessment in support of air quality and climate change governance at different spatial and temporal scales and taking into account the specific regulatory context.[181]

Regarding ethical and societal aspects, questions relating FAIR data are addressed in D9.1.[182] It is also stated that agent-based modelling is used to capture the interactions of population subgroups, industries and service providers in response to the policies considered in the project. Thus, social and cultural factors, socio-economic status and societal dynamics are explicitly taken into account to assess overall policy impact.[183] There also is a work package on ethics requirements, but all deliverables are confidential.[184]

The 3-year project started in January 2018 and will end in December 2020.

The **BodyPass**[185] (API-ecosystem for cross-sectorial exchange of 3D personal data) project aims to revolutionize the health sector and the consumer goods sector breaking down barriers between current data silos. The main incentive behind BodyPass is to facilitate exchange and re-use of 3D data sets from the two sectors.

Based on its dissemination material BodyPass facilitates the secure 3D data exchange and privacy-preserving methods for sharing and aggregating 3D images. [186]

BodyPass is examined more thoroughly in section 3.1.1 in terms of the privacy-preserving technologies used and the assessments tools considered for evaluating the impact of the project's technologies.

The 3-year project started in January 2018 and will end in December 2020.

The **FANDANGO**[187] (FAke News discovery and propagation from big Data ANalysis and artificial intelliGence Operations) project aims to aggregate and verify different typologies of news data, media sources, social media, open data, so as to detect fake news and provide a more efficient and verified communication for all European citizens. The project is validated and tested in three specific domains: climate, immigration and European context, as these are typical scenarios where fake news can influence perception with respect to social and business actions and where news can be verified and validated by trustable information, based on facts and data.

---

[181] https://icarus2020.eu/project-overview/
[182] D9.1 - Data Management Plan, 2018-02, https://icarus2020.eu/wp-content/uploads/2018/03/ICARUS-Deliverable-D9.1_FINAL_REVISED.pdf
[183] https://icarus2020.eu/welcome-letter/
[184] https://icarus2020.eu/work-package/ethics-requirements/#1469709836712-92f0381f-5532
[185] http://www.bodypass.eu/
[186] www.bodypass.eu/sites/default/files/bodypass/public/content-files/article/BodyPass_brochure_online_0.pdf
[187] https://fandango-project.eu/

There are no public deliverables yet, so information on ethical and societal aspects is not available.

The 3-year project started in January 2018 and will end in December 2020.

The **Cross-CCP**[188] (Ecosystem for Services based on integrated Cross-sectorial Data Streams from multiple Cyber Physical Products and Open Data Sources) project aims to establish an IT environment offering data streams coming from mass products, such as vehicles and smart building automation systems. Such data streams lead to the development of new cross-sectorial services, as well as the enhancement of diverse existing services, such as improvement of weather forecast or energy optimisation services.

There are no deliverables yet, but in its objectives it is stated that regarding commercial confidentiality, privacy, IPR and ethical aspects of multiple data streams, the owners can fully control which data they want to provide to whom.[189] It also makes use of an own security, Pprivacy, and trust framework assuring optimal management of confidentiality, privacy, IPR, and ethical issues.[190]

The 3-year project started in December 2017 and will end in November 2020.

The **EDI**[191] (European Data Incubator) project is an 8-month incubation program for big data start-ups and SMEs in Europe that is organised in three phases. There is no information available on ethics or societal aspects and no deliverables have been published yet.

The 3.5-year project started in January 2018 and will end in June 2021.

ICT-15-2016-2017 IAs

In this section, we focus on the IAs funded under ICT-15-2016-2017 (Big data PPP: Large Scale Pilot actions in sectors best benefitting from data-driven innovation). The IAs are tasked to carry out large scale sectorial demonstrations. Again, particular attention is paid to the projects that are in their final year: TT and DataBio. Moreover, we analysed the projects BigMedilytics and BOOST 4.0, which are in their second year. All projects funded under ICT-15-2016-2017 develop big data solutions.

Transforming Transport

The **Transforming Transport**[192] (TT) project aims to demonstrate the transformations that big data brings to the mobility and logistics market in terms of, for instance, increased operational efficiency, improved customer experience and new business models. The project focuses on how big data reshapes transport processes and services using pilots from seven domains that are particularly important for the sector. These domains include smart highways, sustainable vehicle fleets, proactive rail infrastructures,

[188] https://cross-cpp.eu/
[189] https://cross-cpp.eu/objectives/
[190] Cross-CPP Flyer, https://cross-cpp.eu/wp-content/uploads/2018/06/Cross-CPP_Flyer.pdf
[191] https://edincubator.eu/
[192] https://transformingtransport.eu/

ports as intelligent logistics hubs, efficient air transport, multi-modal urban mobility and dynamic supply chains. The project develops various individual pilots.[193]

Apart from the deliverables that focus on general questions such as D1.3, which discusses the project's IPR and data management approach and refers to anonymisation of personal data as one possible solution and D3.2, which describes the project's data management plan and discusses if ethical aspects need to be considered, the project has also released many deliverables on the individual pilots.[194] These deliverables describe the respective pilot's objective, structure and design and also address the data assets used as well as technical aspects of the pilot deployment. With respect to ethical aspects, the deliverables mostly focus on the question if personal data is used or not and, if personal data is used, how the data subjects' privacy is protected.

The 2.5-year project started in January 2017 and will end in June 2019.

DataBio

The **DataBio**[195] (Data-Driven Bioeconomy) project aims to demonstrate the benefits of big data technologies in terms of sustainable improvement and productivity of bioeconomy industry raw materials. For this purpose, a data platform is built based on existing software components that is suitable for different industries and user profiles to ensure effective utilisation of existing datasets, effective participation of the ICT industry and easy setup of new multivendor applications. The project deploys pilots in the fields of agriculture, forestry and fishery. Among the main areas to be addressed by the platform are data acquisition and curation, data variety management, predictive analytics and machine learning, real-time analytics and stream processing, and advanced visualisation and customised technological feedback to the user.

Privacy aspects are addressed in the context of data acquisition and curation. In this regard, D6.2 shortly addresses ethical and privacy issues in DataBio. Furthermore, the project has released public deliverables focusing on the individual pilots. According to them, data privacy and ownership are considered essential elements.[196]

The 3-year project started in January 2017 and will end in December 2019.

Other projects

The **BigMedilytics**[197] (Big Data for Medical Analytics) project has the goal to improve patient outcomes and productivity in the health sector by relying upon big data technologies and using analytics on complex datasets but at the same time safeguarding the security and privacy of personal data. The

[193] https://transformingtransport.eu/transport-domains
[194] https://transformingtransport.eu/downloads/deliverables
[195] https://www.databio.eu/
[196] https://www.databio.eu/wp-content/uploads/2017/05/DataBio_D4.3-Data-sets-formats-and-models_public-version.pdf, https://www.databio.eu/wp-content/uploads/2017/05/DataBio_D4.1-Platform-and-Interfaces_v1.0_2018-05-31_VTT.pdf

project implements twelve pilots that cover three themes with great impact on the sector. The themes covered by the project are population health and chronic disease management, oncology and industrialisation of healthcare services.

A more comprehensive assessment of the BigMedilytics project is provided in section 3.1.4.

The 38-month project started in January 2018 and will end in February 2021.

The **BOOST 4.0**[198] (Big Data Value Spaces for COmpetitiveness of European COnnected Smart FacTories 4.0) project is the biggest European initiative in big data for Industry 4.0. It will lead the construction of the European Industrial Data Space to improve the competitiveness of Industry 4.0 and will guide the European manufacturing industry in the introduction of big data in the factory, providing the industrial sector with the necessary tools to obtain the maximum benefit of big data.

The project has several objectives:

- **Global standards:** Contribution to the international standardisation of European Industrial Data Space data models and open interfaces aligned with the European Reference Architectural Model Industry 4.0 (RAMI 4.0).
- **Secure digital infrastructures:** Adaptation and extension of cloud and edge digital infra-structures to ensure high performance operation of the European Industrial Data Space; i.e., support of high-speed processing and analysis of huge and very heterogeneous industrial data sources.
- **Trusted big data middleware:** Integration of the four main open source European initiatives (Industrial Data Space, FIWARE, Hyperledger, Big Data Europe) to support the development of open connectors and big data middleware with native blockchain support in the European Industrial Data Space.
- **Digital manufacturing platforms:** Open interfaces for the development of big data pipelines for advanced analysis services and data visualisation supported by the main digital engineering, simulation, operations and industrial quality control platforms.
- **Certification:** European certification program of equipment, infrastructures, platforms and big data services for their operation in the European Industrial Data Space.[199]

However, there are no deliverables published yet and there is no information on the consideration of ethical and societal aspects.

The 3-year project started in January 2018 and will end in December 2020.

[198] https://boost40.eu/
[199] https://boost40.eu/objectives/

ICT-16-2017 RIAs

In this section, we focus on the RIAs funded under ICT-16-2017 (Big data PPP: research addressing main technology challenges of the data economy). The RIAs are tasked to address the main technology challenges of the data economy. We analysed the projects CLASS, E2Data, I-BiDaaS, BigDataStack, BigDataGrapes, TYPHON and Track & Know. All projects funded under ICT-16-2017 develop big data solutions and are in their second year.

The **CLASS**[200] (Edge and CLoud Computation: A Highly Distributed Software Architecture for Big Data AnalyticS) project develops a software architecture framework to efficiently distribute big data workloads along the compute continuum. This is done in a completely transparent way, while providing sound real-time guarantees on end-to-end data analytics responses. The capabilities of the CLASS project will be demonstrated on a real smart city use case in Modena with a heavy sensor infrastructure and three connected vehicles equipped with heterogeneous sensors/actuators. The CLASS software architecture will "coordinate edge and cloud computing resources, distribute big-data workloads with real-time requirements along the compute continuum, combine data-in-motion and data-at rest analytics and increase productivity in terms of programmability, portability/scalability and (guaranteed) performance"[201]. Societal and ethical aspects as well as privacy-preserving technologies do not seem to receive a lot of attention in the CLASS project as they are neither mentioned on the website nor in project reports.

The 3-year project started in January 2018 and will end in December 2020.

The **E2Data**[202] (European Extreme Performing Big Data Stacks) project aims to answer two key questions, "How can we improve execution times while using less hardware resources?" and "How can the user establish for each particular business scenario which is the highest performing and cheapest hardware configuration?"[203]. To gain answers to these questions, E2Data proposes an end-to-end solution for big data deployments that improves current infrastructure services. E2Data will test its solutions in four resource-demanding applications: finance, healthcare, green architecture, and security.

E2Data currently works on TornadoVM, a heterogenous virtual machine, which can be used on different devices like CPUs, GPUs, FPGAs and others. This will remove the need for developers to create specific codes for each device.

E2Data has appointed an independent ethics advisor and a data protection officer. Further questions on ethical and societal aspects are not discussed on the project website or in publicly available reports. The 3-year project started in January 2018 and will end in December 2020.

The **I-BiDaaS**[204] (Industrial-Driven Big Data as a Self-Service Solution) project aims to empower IT and non-IT big data experts to easily utilise and interact with big data technologies. I-BiDaaS develops a

[200] https://class-project.eu/
[201] CLASS Brochure: https://class-project.eu/sites/default/files/media/files/brochure_class_print_no_car.pdf
[202] https://e2data.eu/
[203] https://e2data.eu/about/summary
[204] https://www.ibidaas.eu/

solution that will increase the speed of data analysis and facilitates cross-domain data flow. The project develops its own I-BiDaaS architecture, that includes the I-BiDaaS user interface and I-BiDaaS layered system. The user interface is built to address the needs of different user types by different levels of abstractions in programming API, domain language and predefined analytics. The I-BiDaaS architecture has three principal layers: the infrastructure layer, the distributed large-scale layer and the application layer.

Questions of privacy, data protection and data anonymisation are part of the I-BiDaaS architecture and are discussed in the positioning of I-BiDaaS, but do not seem to be a key issue yet.[205] The 3-year project started in January 2018 and will end in December 2020.

The **BigDataStack**[206] (High-performance data-centric stack for big data applications and operations) project aims to provide a fully efficient and optimised cluster management for data operations and data-intensive applications. To do this, the project develops prototypes demonstrating a complete, high-performing, data-centric stack of technologies.[207] BigDataStack develops a data toolkit that has a data-driven infrastructure management system for infrastructure providers and sees Data as a Service for data providers decision makers and private/public organisations.[208] The data toolkit will consist of a process modelling and optimisation framework for business analysts and an application dimensioning workbench for application providers and engineers. BigDataStack works in three industry use cases, real-time ship management, the connected consumer and smart insurance, where the project aims to improve current conditions.

Ethical and societal issues are not a key issue, but the protection of security and privacy is important for the project. BigDataStack aims to propose an architecture that enables security and privacy aspects and that is oriented toward the compliance with data protection regulations.[209]

The 3-year project started in January 2018 and will end in December 2020.

The **BigDataGrapes**[210] (Big Data to Enable Global Disruption of the Grapevine-powered Industries) project aims to improve the production in the wine and natural cosmetics industries by real-time and cross-stream analysis of data sources. BigDataGrapes sees the European wine and natural cosmetics industries in need of big data for being competitive in their market. Therefore, the project develops powerful data processing technologies to support companies in making important business decisions. To demonstrate the developed technologies, BigDataGrapes has established four pilots: the farm management pilot, the natural cosmetics pilot, the table and wine grapes pilot and the wine making pilot.

---

[205] D1.3: Positioning of I-BiDaaS, 2018-09-21, www.ibidaas.eu/sites/default/files/docs/Ibidaas-d1.3%20(1).pdf
[206] http://bigdatastack.eu/
[207] http://www.bigdatastack.eu/project
[208] BigDataStack project presentation, www.bigdatastack.eu/sites/default/files/BigDataStack_Overview_0.pdf
[209] D2.1 – State of the art and Requirements analysis – I, 2018-07-12, www.bigdatastack.eu/sites/default/files/BigDataStack_D2.1_v1.0.pdf
[210] http://bigdatagrapes.eu/

Since BigDataGrapes does not use sensitive data, social and ethical issues in using big data are not a key issue. In case that sensitive data is used, the project plans to anonymise the data before the processing stage and if there is still sensitive data after the processing stage, access controls will be enforced and described.[211] The 3-year project started in January 2018 and will end in December 2020.

The **TYPHON**[212] (Polyglot and Hybrid Persistence Architectures for Big Data Analytics) project aims to provide an industry-validated methodology and integrated technical offering for data persistence architectures that meet the growing scalability and heterogeneity requirements of the European industry. The project deals with use cases in domains such as automotive, earth observation, banking and motorway operation. TYPHON does not only work on a modular event-based architecture for hybrid polystores but also on suitable modelling and query languages. Moreover, the project provides a tool-supported methodology for polystore evolution that takes the presence of multiple, heterogeneous, interdependent and possibly overlapping data stores into account.

Ethical and societal aspects do not seem to have received particular attention yet. The 3-year project started in January 2018 and will end in December 2020.

The **Track & Know**[213] (Big Data for Mobility Tracking Knowledge Extraction in Urban Areas) project seeks to increase the efficiency of big data in the transport, mobility, motor insurance and health sectors by creating a scalable, fault-tolerant platform for big data. The project suggests a multi-disciplinary approach considering the needs of researchers, customers and the producers and providers of services. Track & Know focusses its development of software and efficient, interoperable and scalable toolboxes on automotive transportation and urban mobility in general. The toolboxes in development are: Data Processing Architectures & Infrastructure (BDMI Toolbox), Big Data Processing Toolboxes Management (BDP Toolbox), Big Data Analytics (BDA Toolbox) and Data Visualization & User Interaction (VA Toolbox).

Ethical and societal aspects do not seem to be a main issue in the project yet. The 3-year project started in January 2018 and will end in December 2020.

ICT-13-2018-2019 RIAs and CSA

In this section, we focus on the RIAs as well as the CSA funded under ICT-13-2018-2019 (Supporting the emergence of data markets and the data economy). The projects are tasked to support the emergence of data markets and the data economy. We analysed the RIAs Safe-DEED, MOSAICrOWN and MUSKETEER as well as the CSA Data Market Services. While Data Market Services supports the development and use of big data solutions, the other projects develop big data solutions. All projects funded under ICT-13-2018-2019 are in their first year.

The **Safe-DEED**[214] (Safe Data Enables Economic Development) project aims to improve security technologies and build trust towards privacy-preserving technologies by bringing together cryptography,

[211] D2.2 Data Management Plan & Support pack, 2018-03-30, https://zenodo.org/record/1319279#.XH5A2qBCfRa, p. 13
[212] https://www.typhon-project.org/
[213] https://trackandknowproject.eu/
[214] https://safe-deed.eu/

data science, business innovation and the legal domain. Furthermore, Safe-DEED aims to provide a set of tools to facilitate the assessment of data value. To achieve these goals, the project develops new methods addressing the problems of private set intersection for large data sets, and secure multi-party computation (MPC), works on a Software as a Service (SaaS) component for valuating the potential of a company's data if collected and processed at scale, and shows the scalability of the methods in real-world scenarios.[215] On top of that, Safe-DEED will show the general applicability and sustainability of the methods for MPC and data valuation, develop pricing models and business models and try to generate trust in data markets.[216]

Ethical and societal issues play an important in role in the project. The project addresses how the economy can profit from privacy-preserving technologies[217], what legal and ethical issues can occur[218] and what the knowledge value of a certain corpus of a structured data is without having to completely analyse it[219]. Moreover, Safe-DEED works on protocols for privacy, confidentiality and de-anonymisation.[220]

The 3-year project started in December 2018 and will end in November 2021.

The **MOSAICrOWN**[221] (Multi-Owner data Sharing for Analytics and Integration respecting Confidentiality and Owner control) project "aims to enable data sharing and collaborative analytics in multi-owner scenarios in a privacy-preserving way, ensuring proper protection of private/sensitive/confidential information."[222] The goal is to allow data owners to maintain control over their own data during the data sharing process by providing efficient and scalable privacy-aware collaborative computations. To achieve this goal, the project works on a data governance framework, and effective and efficient protection techniques. Therefore, efforts are made to support a rich set of protection requirements, to provide a data governance framework, to provide efficient and effective techniques for data wrapping, to provide efficient and effective techniques for data sanitisation and to provide effective exploitation in real operational environments.[223]

MOSAICrOWN has a work package that deals with ethical requirements. It has not yet produced any public deliverables. The project's approach to privacy preservation will be closely watched in the future. The 3-year project started in January 2019 and will end in December 2021.

The **MUSKETEER**[224] (Machine learning to augment shared knowledge in federated privacy-preserving scenarios) project aims to create a validated, federated, privacy-preserving machine learning platform that is interoperable, scalable and efficient enough to be deployed in real use cases. The project wants

[215] https://safe-deed.eu/project-partners/
[216] https://safe-deed.eu/project-partners/
[217] https://safe-deed.eu/wp-business-model-innovation/
[218] https://safe-deed.eu/wp-legal-ethical/
[219] https://safe-deed.eu/wp-data-value/
[220] https://safe-deed.eu/wp-security-privacy-2/
[221] https://mosaicrown.eu/
[222] Ebd.
[223] https://mosaicrown.eu/the-project/objectives/
[224] http://musketeer.eu/

to provide secure, scalable and privacy-preserving analytics over decentralised datasets using machine learning. The advantage will be that data can still be stored in different locations with different privacy constraints but shared securely.[225] The goals of MUSKETEER are to create machine learning models over a variety of privacy-preserving scenarios, to ensure security and robustness against external and internal threats, to provide a standardised and extendable architecture, to demonstrate and validate in two different industrial scenarios and to enhance the data economy by boosting sharing across domains.[226]

Since MUSKETEER uses a privacy-preserving approach to using data, ethical and societal questions will be addressed within the scope of the project. The 3-year project started in December 2018 and will end in November 2021.

The **Data Market Services**[227] (Supporting the European data market providing free support services to data-centric SMEs and start-ups) project will support the European data market by providing free support services to data-centric SMEs and start-ups. The services offered by the project are related to fundraising (public fundraising and venture capital match-making), acceleration (entrepreneurial trainings, acceleration and incubation mobility programme and mentoring), standards and legal aspects (IPR for entrepreneurs, GDPR training and standards awareness), promotion (marketing for startups and TNW package), and data skills (data science academy and trust building tools). These services will be offered in the form of webinars, live trainings, contents, mentorship, a mobility programme and promotion.

It is not yet foreseeable how and to what extent ethical and societal issues will be addressed. The 3-year project started in January 2019 and will end in December 2021.

Other projects

Apart from the projects discussed in section 2.1 to 2.5, the projects BDVe and DataBench, which are funded under ICT-17-2016-2017 (Big data PPP: Support, industrial skills, benchmarking and evaluation), are particularly relevant for e-SIDES. Additionally, the projects CANVAS, K-PLEX, CLARUS and BYTE were analysed. While CANVAS is in its final year, the other three projects have already ended in 2018 and 2017, respectively. The projects have received funding under different H2020 and FP7 programmes and topics.

The **BDVe**[228] (Big Data Value ecosystem) project provides coordination and support to the Big Data Value Public-Private Partnership (PPP) projects funded by the EU. To ensure this, BDVe goals are to achieve a more competitive landscape of European data providers, to create the context for a more competitive EU industry and to ensure the sustainability of the investments and actions triggered by the PPP. The project's priorities are therefore to be accurately informed about the most important facts in big data, to support the implementation of the big data PPP from an operational point of view, to develop a vibrant community around the PPP, to support the development of a European network of

[225] http://musketeer.eu/project/
[226] Ebd.
[227] https://www.datamarketservices.eu/
[228] http://www.big-data-value.eu/

infrastructures and centres of excellence around big data, to set-up a professional communications strategy, to set-up a framework that supports the acceleration of data-driven businesses, and to ensure the sustainability of the investments and actions triggered by the PPP.

To support the projects within the PPP, BDVe has created a the BDV Marketplace, a one-stop shop for platforms, technologies and solutions, where interested people can search the catalogue of innovative solutions on big data, platforms, services and technologies developed under the PPP, participate by asking questions and finding answers related to the different solutions available in this marketplace and contribute by uploading information about solutions from their projects and make them reachable to the BDV community through this marketplace.[229] BDVe also provides i-Spaces, Trusted Data Incubators targeted to accelerate take-up of data-driven innovation in commercial and non-profit sectors.[230] Another task of BDVe is the support of SMEs, where a group of small and medium-sized enterprises has been selected as active contributors to the development of the big data ecosystem, to unlock the potential of a European data economy.[231] BDVe will also provide a landscape, a map to show how and where industry, academia, public sector and other actors are working to address the societal challenges of the usage of big data, and several webinars, where results of EU projects funded under topics related to the Big Data Value PPP are presented.

The 4-year project started in January 2017 and will end in December 2020.

The **DataBench**[232] (Evidence Based Big Data Benchmarking to Improve Business Performance) project has the goal to design a benchmarking process helping organisations developing big data technologies, by measuring their technology development activity against parameters of high business relevance. DataBench investigates existing big data benchmarking tools and projects, identifies the main gaps, and provides a set of metrics to compare technical results coming from those tools. The project develops a framework to associate the technical results with the economic processes that are imperative to a company. Another part of DataBench is a toolbox that provides a unique environment to search, select and deploy big data benchmarking tools. The project will also publish a handbook that provides guidelines to the use of the project's results, framework an toolbox, describing metrics implementation and benchmarks.

Ethical and societal issues are not yet a recognizable part of DataBench.

The 3-year project started in January 2018 and will end in December 2020.

The **CANVAS**[233] (Constructing an Alliance for Value-driven Cybersecurity) project aims to construct a community that unifies the perspectives of legal and philosophical scholars with empirical researchers and makes sure that technology development in cybersecurity incorporates European values and fundamental rights. CANVAS structures existing knowledge, designs a network for exchanging

[229] http://marketplace.big-data-value.eu/
[230] http://www.big-data-value.eu/i-spaces/
[231] http://www.big-data-value.eu/ecosystem/
[232] https://www.databench.eu/
[233] https://canvas-project.eu/

knowledge and generating insights across domains, and disseminates the insights gained through three means: a reference curriculum for value-driven cybersecurity with a focus on industry-training, briefing packages for policy stakeholders, and a MOOC (massive open online course) on value-driven cybersecurity. A CANVAS book on the ethics of cybersecurity will be published in August 2019. The book will discuss the ethical aspects of cybersecurity and have a strong practical focus, including case studies that outline ethical dilemmas in cybersecurity and present guidelines and other measures to tackle those dilemmas.[234]

Questions of ethics and privacy are a main part of the CANVAS material and workshops.

The 3-year project started in September 2016 and will end in August 2019.

The **K-PLEX**[235] (Knowledge Complexity) project investigated the elements of humanities and cultural data, and the strategies researchers have developed to deal with them to better shed light on the gap between analogue or augmented digital practices and fully computational ones. The project's three main challenges were to find out the manner in which data that are not digitised or shared become 'hidden' from aggregation systems, the fact that data is human created and lacks the objectivity often ascribed to the term, and the subtle ways in which data that are complex almost always become simplified before they can be aggregated.

The main outcome of the K-PLEX project is a document entitled "Big Data & Complex Knowledge. Observations and Recommendations for research from the Knowledge Complexity".[236] In this paper big data is discussed by a humanities approach, which includes societal and ethical aspects of big data. The outcomes are:

- Big data is ill-suited to representing complexity: the urge toward easy interrogability can often result in obscurity and user disempowerment.
- Big data compromises rich information.
- Standards are both useful and harmful.
- The appearance of openness can be misleading.
- Research based on big data is overly opportunistic.
- How we talk about big data matters.
- Big data research should be supported by a greater diversity in approaches.
- Even big data research is about narrative, which has implications for how we should observe its objectivity or truth value.

K-PLEX recommends to enhance regulation of big data research, to rethink the disciplines that contribute to big data research, to reverse knowledge hierarchies within big data research to disrupt

[234] https://canvas-project.eu/results/canvas-book.html
[235] https://kplex-project.com/
[236] Edmond, Jennifer et al.: Big Data & Complex Knowledge. Observations and Recommendations for research from the Knowledge Complexity. https://kplexproject.files.wordpress.com/2018/06/trinity-big-data-report-jklr_04-1.pdf

biases and fixed mindsets, and to ensure contextualised data sharing for big data research, keeping context as minimal as necessary, but as rich as possible.

The 15-month project started in January 2017 and ended in March 2018.

The **CLARUS**[237] (A Framework for User Centred Privacy and Security in the Cloud) project aimed to enhance trust in cloud computing services by developing a secure framework for the storage and processing of data in the cloud that allows end users to monitor, audit and retain control of the stored data without impairing the functionality and cost-saving benefits of cloud services. CLARUS provides the end user with a dedicated proxy located in a trusted domain implementing security and privacy features towards the cloud provider. CLARUS also provides a set of security auditing services enabling the user to supervise the security operations performed by the CLARUS framework as well as other trust-enhancing features. These services include a Security Policy Manager, an Access Rights Management tool and a Security Administrator.

The CLARUS project uses several privacy techniques for the proxy and the tools, which include data anonymisation, data encryption, data splitting, homomorphic encryption and searchable encryption.[238] The project also developed some legal guidelines for data quality and data subject rights. The data quality principles consist of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality and the data subject's rights consist of information, access, rectification, erasure, restriction and portability. Furthermore, CLARUS published a deliverable on the legal and ethical requirements of the project.[239] The requirements range from the above mentioned guidelines to general security guidelines, breach notification and contractual mitigation of liability. All the identified requirements are considered a must have, meaning "the project will be seriously impacted if the requirement is not net."[240]

A more comprehensive assessment of the CLARUS project is provided in section 3.1.3.

The 3-year project started in January 2015 and ended in December 2017.

The **BYTE**[241] (Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities) project aimed to assist European science and industry in capturing the positive externalities and diminishing the negative externalities associated with big data. BYTE created an advisory board and additional network contacts to conduct a series of big data case studies in actual big data practices across a range of disciplinary and industrial sectors to gain an understanding of the economic, legal, social, ethical and political externalities that are in evidence. The main outcome of the BYTE project is the BYTE handbook with policy recommendations and research roadmaps.[242] The social and ethical

[237] http://clarussecure.eu/
[238] http://clarussecure.eu/privacy-techniques
[239] D2.4 – Legal and Ethical Requirements, 2017-02-17, http://clarussecure.eu/sites/default/files/CLARUS-D2.4-LegalandEthicalRequirements.pdf
[240] ebd. p. 58
[241] http://new.byte-project.eu/
[242] D7.3 - BYTE Final Report and Guidelines, 2017-02-28, http://new.byte-project.eu/wp-content/uploads/2014/02/D7.3-Final-report-FINAL.pdf

findings of the project are that better services are provided, improved decision-making is possible, fear of discrimination through big data is perceived, inequalities are reproduced by big data, higher citizen participation is hoped for and a general lack of trust is present.[243]

Ethical and societal issues, including privacy, are main part of the roadmaps and the entire project.

The 3-year project started in March 2014 and ended in February 2017.

[243] ebd. p. 4f