Policy Cloud

Cloud for Data-Driven Policy Management

CLOUD FOR DATA-DRIVEN POLICY MANAGEMENT

Project Number: 870675 Start Date of Project: 01/01/2020 Duration: 36 months

D3.3 POLICYCLOUD'S SOCIETAL AND ETHICAL REQUIREMENTS & GUIDELINES

Dissemination Level	PU
Due Date of Deliverable	31/12/2020, M12
Actual Submission Date	29/12/2020
Work Package	WP3 (Cloud Infrastructures Utilization & Data Governance)
Task	3.5
Type	Report
Approval Status	
Version	V1.0
Number of Pages	p.1 – p.122

Abstract: This deliverable analyses the ethical, legal, regulatory, and societal issues related to PolicyCLOUD, also providing a synthetic review of the existing debate and literature on these topics. Also, a specific analysis is dedicated to the issues related to each use case. Finally, guidance is provided on how the ethical, legal, regulatory, and societal requirements shall be embedded in the solutions developed throughout the Project, also identifying a list of controls to be used to continuously monitor the compliance of PolicyCLOUD with the identified requirements.

PolicyCloud has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870675.

Versioning and Contribution History

Version	Date	Reason	Author
0.0	10/12/2020	First draft	A. Audino A. Bettiol A. Strippoli M. Taborda Barata
0.1	17/12/2020	Peer Review	R. Munné Caldes
0.2	21/12/2020	Address peer review comments	A. Audino A. Bettiol A. Strippoli M. Taborda Barata
0.3	21/12/2020	Peer Review	V. Moulos
0.4	22/12/2020	Address peer review comments	A. Audino A. Bettiol A. Strippoli M. Taborda Barata
0.5	22/12/2020	Quality Check	A. Mavrogiorgou
1.0	29/12/2020	Deliverable ready for submission	A. Audino A. Bettiol A. Strippoli M. Taborda Barata

Author List

Organisation	Name
ICTLC	A. Audino
ICTLC	A. Bettiol
ICTLC	A. Strippoli
ICTLC	M. Taborda Barata

Abbreviations and Acronyms

Abbreviation/Acronym	Definition
AI	Artificial Intelligence
AI-HLEG	High-Level Expert Group on Artificial Intelligence
CFREU	Charter of Fundamental Rights of the European Union
CoE	Council of Europe
CNIL	Commission Nationale de l'Informatique et des Libertés
CRC	Convention on the Rights of the Child
CSA	Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)

Abbreviation/Acronym	Definition
CSPCERT	European Cloud Service Provider Certification
DGA	Data Governance Act
DPA	Data Processing Agreement
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DSM	Digital Single Market
EEA	European Economic Area
EC	European Commission
ECHR	European Convention on Human Rights
ECI	European Cloud Initiative
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EESC	European Economic and Social Committee
EGE	European Group on Ethics in Science and New Technologies
ENISA	European Union Agency for Cybersecurity
EOSC	European Open Science Cloud
EU	European Union
FRIA	Fundamental Rights Impact Assessment
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
GTD	Global Terrorism Database
IaaS	Infrastructure as a Service
ICCPR	International Covenant on Civil and Political Rights
ICT	Information and Communication Technology
IE	Information Ethics
IoT	Internet of Things
ISMS	Information Security Management System
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
LIA	Legitimate Interest Assessment
OECD	Organization for Economic Cooperation and Development
PDT	Policy Development Toolkit
SLA	Service Level Agreement
SMEs	Small and Medium-Sized Enterprises
SWIPO	Switching Cloud Providers and Porting Data
TEU	Treaty on the European Union
TFEU	Treaty on the Functioning of the European Union
UN	United Nations
USA	United States of America
WP	Work Package
WP29	Article 29 Data Protection Working Party

Contents

List of Tables

Executive Summary

This deliverable analyses the ethical, legal, regulatory, and societal issues related to PolicyCLOUD, also providing a synthetic review of the existing debate and literature.

With regards to ethical and societal issues, from a general standpoint the main findings relate to the importance of ensuring the accuracy of the dataset used for performing the analytics and the policymaking to achieve an adequate degree of reliability on the policies developed on the basis of the same analytics. Also, the respect of the principle of transparency appears relevant to ensure the engagement of the end-users and to obtain their trust in the policies developed through PolicyCLOUD. Moreover, the key issue is to ensure an adequate level of human engagement in the data processing and policymaking processes, to avoid the relevant ethical and societal risks related to a complete automatization of decisional processes, which may be jeopardised by biases (whether in the initial dataset or in the algorithm), leading for example to discrimination phenomena.

The general legal and regulatory issues related to the Project concern contractual protection of data sources, legal protection of databases, copyright, and personal data protection and privacy. Of this list, personal data protection is the most important legal and regulatory issue related to the Project, since the development of PolicyCLOUD implies the collection and processing of a relevant amount of personal identifiable information. Therefore, the compliance with the requirements defined by the GDPR and other applicable personal data protection regulations is paramount for the correct and sustainable implementation of PolicyCLOUD.

Also, by analysing in detail the ethical, legal, regulatory, and societal issues related to the components of the Project, the risks appear to be focused on the selection of the datasets to be used, from the perspective of both their accuracy and the legitimacy to collect and process the data for the purposes of the Project. These issues need to be addressed whether the data used constitute personal identifiable information; however, when personal data are involved, appropriate safeguards shall be implemented, especially to comply to applicable data protection laws.

The exam of the specific issues related to each of the use case highlights the risks associated with the first use case, since the related activities can create interferences with some of the fundamental rights recognized by the CFREU, the ECHR, and other international legal instruments, not to mention the common constitutional tradition of the EU member State. Therefore, it will be of the utmost importance to monitor the development of the activities related to this use case, to ensure their compliance to the applicable ethical, legal, regulatory, and societal requirements. Also, for the other use cases, some specific issues have been identified, mostly related to personal data protection.

At the end of the deliverable, some guidance is provided on how the ethical, legal, regulatory and societal requirements shall be embedded in the solutions developed throughout the Project, also identifying a list of controls to be used to continuously monitor the compliance of PolicyCLOUD with the identified requirements.

1 Introduction

The PolicyCLOUD project ("**PolicyCLOUD**" or the "**Project**") aims to harness the potential of digitization, big data, and cloud technologies to improve the modelling, creation, and implementation of policies. In three years (2020-2022) PolicyCLOUD will address challenges faced by many businesses and public administrations of improving how they make policy decisions by accessing and using data.

The aim of PolicyCLOUD is to deliver a unique, integrated environment of curated datasets and data management, manipulation, and analysis tools addressing the full lifecycle of policy management in four distinct thematic areas, and using the data analysis capabilities of the ECI, with an emphasis on data analysis to facilitate evidence-based policy making. PolicyCLOUD introduces a pioneering approach for the development of policies collections to exploit collective knowledge towards policy co-creation and cross-sector optimization.[1]

To maximize societal acceptability and trust in PolicyCLOUD, and through it, in policies, the PolicyCLOUD consortium (the "**Consortium**") is aware of the necessity of providing extensive and in-depth analyses of legal, regulatory, societal and ethical aspects, by seeing to an optimal embedding of the results of these into the design of the solution, and by thoroughly evaluating the extent to which this has been successful. Special attention must be paid on the ethical and societal issues which will need attention throughout the Project. Therefore, it is necessary to identify a set of system dimensions, features, and functionalities, and their links to the range of socially and ethically significant new practices that the system enables, and to propose a refined set of requirements guidelines and norms for the responsible modelling of policies, aligned with the iterations of the development and demonstration of the platform in the use cases.

This deliverable will analyse the ethical, legal, regulatory, and societal issues related to digitization, big data, and technologies in general, also providing a synthetic review of the existing debate and literature on these topics. Then, the specific issues related to each of the use cases will be examined. Finally, we will provide some guidance on how the ethical, legal, regulatory and societal requirements shall be embedded in the solutions developed throughout the Project, also identifying a list of controls to be used to continuously monitor the compliance of PolicyCLOUD with the identified requirements.[2]

This document will be updated during the Project at M22 and M34.

[1] For more information about the Project, see https://cordis.europa.eu/project/id/870675 and https://policycloud.eu/.

[2] With regards to legal and regulatory issues, the scope of the analysis, in the context of this deliverable, will be generally limited to EU and international law, without exploring in detail the specific national and/or local requirements related to the countries and jurisdictions in which the use cases are implemented. Nevertheless, where specific analysis on local and/or national regulations shall result as appropriate and/or necessary, we will highlight this as a field for which further research is needed and that will be consequently developed in the next versions of this deliverable to be released at M22 and M34.

2 General ethical, societal, legal, and regulatory issues

2.1 Ethical and societal issues

2.1.1 Introduction

The current global shift to an ever more interconnected reality has caused a corresponding shift in the way public authorities, organisations and people conceive the world around them.[3] There is an observable societal trend in transitioning from physical, on-paper tools, products, and activities to their more effective and technologically advanced digital counterparts. In sectors and activities where this transition has not yet been fully achieved (e.g., the definition of public policy within Europe), the demand for the ability to harness the computational power of cloud-based systems, capable of processing large amounts of information through meticulously crafted algorithms, is growing. [1]

This shift brings about a need to change the ways in which information is collected, used, and managed through technologies such as cloud computing, big data processing and AI. The gains in efficiency and effectiveness brought about by these technologies must be balanced against the potential impact which they may have on the fundamental rights and freedoms of individuals, as well as on the functioning of society at large. This is important:

1. From a value system perspective, to preserve the European commitments reflected, among other legal instruments, in the CFREU.

2. From a practical perspective, so as to maximize the trust of the citizens in these technologies and their acceptance by the same citizens, to minimize the harmful impacts which they may have on society and to appropriately harness their benefits so as to improve the overall quality of life for humanity. Indeed, many studies on the adoption of new technologies show how the impact of the technology would change whether there is an adhesion to the ethical principles from the beginning or not. [1] [3]

As noted by the EC, [4] the European approach towards the use of new technologies and their impact on the rights of individuals should be contrasted to those of other countries, such as the models followed by the USA and China:

1. In the USA, the organization of the data space is left to the private sector, with considerable concentration effects. Indeed, economic research seems to have reach a substantial consensus on the fact that the market share reached by the American big tech corporations is jeopardising the dynamics of free market, also finding that one of the factors that has facilitated the current concentration levels is the absence in the USA of effective personal data protection regulations. [5]

[3] As examples of the impact of these shifts, public authorities are beginning to adopt big data and cloud-based solutions in their governmental and regulatory activities, organisations are progressively increasing their reliance on interconnected services and open data, and individuals themselves consume larger and larger amounts of online content.

2. On the other hand, China relies on a combination of government surveillance with a strong control of big tech companies over massive amounts of data, without sufficient safeguards for individuals.

The EC underlined the importance for Europe "[…] *to find our European way, balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards*".[4] [4]

To remain aligned with the approach advocated by the EC, the design and operation of the PolicyCLOUD platform, which leverages all three of the mentioned technologies, must therefore be assessed not only from a legal and regulatory perspective to ensure its lawfulness, but also from an ethical and societal perspective to ensure its alignment with ethical and societal values.

In this section, we will explore the existing literature on ethical and societal issues of relevance to the Project, including issues related to use of cloud systems, big data and AI (e.g., liability of AI systems, risks related to algorithm biases, engagement of citizens in the policymaking process, risks related to concentration of technological instruments governance on specific social groups, etc.), also considering statements and guidelines issued on this topic by relevant European and international institutions. The aim of this section is to identify a set of ethical and societal requirements which can be embedded in the solutions developed in the context of PolicyCLOUD.

2.1.2 General ethical and societal considerations on cloud computing

The world today is increasingly cloud-enabled, and the implications for citizens and organisations are far-reaching. The cloud offers new ways to access data and an expanded computing capacity to process it quickly and inexpensively, boosting innovation and assisting organisations in developing new and disruptive technologies and practices.[5] However, while the effect of the cloud within the technology world may be relatively clear, its impact on society is currently just starting to be understood: therefore, decision makers have a responsibility to recognize the implications and help ensure they are addressed. [6]

As a result of the enhanced capabilities offered to organisations through cloud computing, mass amounts of data, whether personal data or not, can now be amassed into central cloud repositories, to be analysed by meticulously crafted algorithms to allow the identification of heretofore unknown or even undetectable patterns and trends in those data. The incentive offered by cloud computing (to maximize the collection of data and, thereby, the value which can be extracted from it) may, if not properly checked, present security risks (e.g., related to new types of vulnerabilities created by use of geographically distributed infrastructure and data storage) and risks to the privacy, the protection of personal data and other fundamental rights of individuals (e.g., related to the collection of excessive personal data and the use of those data in violation of the reasonable expectations of individuals). These risks are joined by those created using AI, including machine learning systems, which is often deployed through cloud-based systems to maximise the ability to extract knowledge from big data pools.

Cloud computing is at the centre of European discussions and initiatives, such as:

[4] The EU has already taken several steps since 2014. Indeed, with the GDPR, the EU has created a solid framework for digital trust. Other initiatives, such as the regulation on the free flow of non-personal data (Regulation (EU) 2018/1807), the CSA, and the Open Data Directive (Directive (EU) 2019/1024) are all inspired by this vision.

[5] DevOps practices, the IoT, AI, edge computing and cryptocurrency are all examples which came about and have developed thanks, in general, to the scalability, efficiency and ubiquity of cloud computing.

1. The Regulation (EU) 2018/1807 ("Regulation on the free flow of non-personal data"), that, together with the GDPR, raises legal certainty for cloud users, by ensuring the free movement of all data in the EU.

2. The self-regulatory work on cloud switching and cloud security addressed by the DSM Cloud Stakeholder Groups, which has resulted in the recently finalized SWIPO data portability codes of conduct and the CSPCERT recommendations for a candidate European cloud security certification scheme. [7][8] Furthermore, at the request of the EC, the ENISA is working on a single European cybersecurity certification scheme for cloud services. [9] The scheme will provide increased assurance to businesses, public administrations, and citizens that their data is secure wherever they are stored or processed. The EC has also facilitated a platform for the industry to develop codes of conduct related to data protection in cloud environments. This has resulted in two codes of conduct that are currently being reviewed by the EDPB. [9]

In the context of the cloud, the notions of data protection and intellectual property rights, as well as data access, ownership, and control, among others, need to be reinterpreted and given new meaning. The decentralized nature of cloud computing systems increases the exposure of data processed through those systems to interception, loss, alteration, or misuse. The often opaque and complex processing chains involved in the operation of these systems, with potentially multiple different actors providing the data collection, storage, analysis and manipulation functionalities for a single system from geographically diverse locations, makes it difficult for individuals to be aware as to where and how their information is processed on the cloud, and therefore also to exercise control over their own information.

The ethical issues related to the use of cloud computing and big data are explored by the IE discipline, a branch of applied ethics focused on the "*production, dissemination, storage, retrieval, security, and application of information within an ethical context*". [10] In the IE context, core ethical issues related to cloud-based solutions that have been pointed out are related to confidentiality, privacy and personal data protection, the presence of biases in the information provided, the quality of data supplied and the use of work facilities.

Further key ethical concepts to be considered are those of responsibility and trust.

1. *Responsibility*. Ethical behaviour of one person toward another implies a relationship of responsibility in which there is a subject, the entity or agent held responsible, and an object, the entity or agent to which the subject bears some responsibility. This relationship of responsibility needs to be supported by norms of behaviour, and by the mechanisms that establish the relationship and maintain it so that it is workable: however, the identification of the elements of this relationship in the cloud context appears to be a problem not easy to solve. [12]

2. *Trust*. Trust may be defined as the confidence of one party in another with respect to specific actions or benefits and involving a relationship of voluntary vulnerability, dependence, and reliance. In the realm of cloud computing, a relationship of trust must be developed between cloud service providers, cloud service customers/users and the individuals affected by use of the cloud service, where these are not the customers/users themselves. One mechanism through which trust can be generated is the assurances to which the cloud service provider is willing to bind itself in the service contract entered with its customers/users. [13]

From an ethical standpoint, it would be useful to reflect on minimum standards to be granted in contracts to be signed with cloud service providers, to re-balance the position of customers and users. [14] Governments, when leveraging cloud-based systems to process information on their citizens, must understand how citizens may feel about sharing their data with commercial bodies (i.e., cloud service providers), and how to reinforce the trust and confidence in, as well as acceptance of, such cloud-based systems from the same citizens.

To allow the respect of these requirements and given that the cloud-based infrastructure on which the PolicyCLOUD platform relies is supported by an external IaaS provider, it is important to have a clear and specific framework in place with the IaaS provider, in which objectives, processes and results expected from PolicyCLOUD platform are clearly set out, so as to specify the capabilities needed from the IaaS provider. A solid framework will provide PolicyCLOUD with greater certainty as to what it can expect, from the technical perspective (e.g., availability, data security, limitations on access to data, geographical location of data storage, etc.), from the cloud-based infrastructure on which the platform is to be hosted and developed, and to further transmit that certainty to the end-users of the same platform (i.e., policy makers). This, in turn, is fundamental for policy makers to be able to justify the choice to rely on a cloud-based solution towards the individuals for which they have responsibility (e.g., citizens under the purview of the policy maker), allowing them to potentially assuage relevant concerns which those individuals may have.

This framework should allow PolicyCLOUD to monitor the commitments assumed by the IaaS provider in this respect, including through specific audits throughout the contractual relationship and even beyond (e.g., to determine whether the IaaS provider has properly abided by its obligations to delete data stored on behalf of PolicyCLOUD). Limitations on the possibility for unilateral amendments, by the IaaS provider, to the capabilities offered are likewise of paramount importance, to allow PolicyCLOUD to react appropriately (by ensuring that amendments do not cause cloud service levels to drop to unsatisfactory levels, or otherwise to consider alternative providers in a timely manner). A substantial change to the cloud capability offerings of the IaaS provider may jeopardise the viability of the platform, impacting all the end-users of the platform and potentially also individuals affected by platform-based policy making activities performed by the same end-users.

For this same reason, one crucial aspect with ethical implications is the potential impact of a potential loss of availability affecting the cloud-based system. Availability, in this sense, relates to the availability of the cloud infrastructure (i.e., the hardware framework used to support the cloud computing system and the cloud-based software which may be run on that framework). The continuity of certain cloud-based services and the possibility to access the data processed through that service at any time may be crucial, depending on the context in which an end-user leverages the PolicyCLOUD platform, particularly where the associated policy making activities are dependent on access to data updated in real time, where an overly long delay in access to data or an interruption of service may significantly impact the ability of the policy maker to accurately configure its policies. This, in turn, may increase the risk of inaccurate or misguided policy making, potentially leading to inequitable, discriminatory, or otherwise maladjusted policy development. The framework developed with the IaaS provider (e.g., defined through a SLA) should therefore define appropriate service levels so as to ensure that the platform and its data will be kept promptly available to PolicyCLOUD and end-users, identifying a maximum amount of acceptable service downtime and ensuring the possibility to recover data which may be lost during the interruption. Infringement of these levels should preferably be subjected to appropriate contractual penalties.

Furthermore, the issue of vendor lock-in must be appropriately addressed with the IaaS provider. Should it be necessary or adequate for PolicyCLOUD to engage an alternative provider, continuity of service and data stored on the platform should be ensured during the transition.

From a societal perspective, it is important also to consider the potential impact on environmental sustainability which leveraging a cloud-based infrastructure may have, as opposed to relying on local infrastructure and/or storage capabilities. It is very difficult to assess the environmental aspects that come with cloud computing and to understand if it can lead to less emissions and energy consumption given the reliance on shared resources, or if those shared resource will produce even more emissions. PolicyCLOUD should ensure that reliance on a cloud-based system is not only a more efficient solution in terms of the technical capabilities offered as opposed to non-cloud-based alternatives, but also that the environmental impact of the infrastructure necessary to support the functioning of that system is reduced to a minimum, so that the benefits are not ultimately outweighed by this impact over time.

2.1.3 General ethical and societal considerations on big data management

Big data management is a relevant topic to be considered, since the PolicyCLOUD platform will use analytics technologies using large datasets to enable policy makers to undertake their decisions.

Today, an enormous amount of data is being continuously generated in all walks of life by all kinds of devices and systems, every day. A significant portion of such data is being captured, stored, aggregated, and analysed in a systematic way without losing its characteristics. [15]

The concept of big data refers to the practice of combining huge volumes of diversely sourced information and analysing them, often using self-learning algorithms to inform decisions.[6] [16] This information is not always personal: data generated by sensors for monitoring natural or atmospheric phenomena like the weather or pollution (as collected by one of the data sources identified as relevant for the third use case of the Project) do not relate to an identified or identifiable natural person.

A useful definition of big data was first provided in 2014 and later modified by IBM scientists and it has become more and more accepted. [17] According to this definition, big data is characterized by:

1. Volume, referring to the scale of data.
2. Variety, since data is produced by different data sources in different formats.
3. Velocity, which is connected to the analysis of streaming data.
4. Veracity, as data is uncertain and needs to be verified before or during use.
5. Value, which can be produced by analysing big data.

Though not all big data is personal data, as not all of it relates to individual human beings, one of the greatest values of big data for both private organisations and governments derives from the monitoring of human behaviour, collectively and individually, and resides in its predictive potential. Given that it is becoming progressively easier to infer the identity of individuals from allegedly anonymous datasets, particularly through combination of those datasets with others, including publicly available information (e.g., data retrieved on social media), it is not a simple matter for big data to be fully and properly anonymised, thereby subjecting such big data to the requirements of data protection law. [18] Where that data is traded especially across borders and jurisdictions, accountability for processing the information becomes nebulous and difficult to ascertain or enforce under data protection law, particularly in the absence of any international standards.

Recent developments in data-driven information systems set big data research and business analytics at the core of computer science and social science. In computer science research, there is a consensus that big data and data analytics research will foster a new generation of information systems capable of managing collective wisdom in human decision making and smart machines. Emerging research areas like cognitive computing, combined with AI and machine learning, permit advanced and sophisticated methods for processing data, including sentiment

[6] In the referenced opinion (see end note), the WP29 states that: "*Big data refers to the exponential growth both in the availability and in the automated use of information: it refers to gigantic digital datasets held by corporations, governments and other large organisations, which are then extensively analysed (hence the name: analytics) using computer algorithms*".

analysis, image processing, natural speech recognition and text mining. In parallel, emerging technologies including cloud computing, IoT and virtual reality allow applications and services to increase their data processing capabilities and provide an expanded and enhanced range of services to their users. The development of a huge data ecosystem around the globe, in which providers and users of data promote business value in terms of data and decision making, is a key development of our times. In this context, users of applications and services worldwide participate consciously, or unintendedly, to an integrated data dissemination and aggregation process with critical trust and personal data protection issues.

The concept that data is shared and stored on servers through the use of the internet implies that this process can take place in two ways, depending on how the interaction occurs between the subject that produces data and the storage system. According to the kind of interaction, we can identify:

1. Active big data, when a user directly sends data to a storage system.[7]

2. Passive big data, when the data of an individual is collected by another person and then input into an online storage system.[8]

Data transfer, in fact, regardless of how direct the connection between a subject and the storage system is, may take place in more or less explicit ways. It is, therefore, possible to distinguish between:

1. Consciously transferred data, when a user is timely and clearly informed that data about him or her is being collected and stored, and therefore awareness can be safely assumed.

2. Not-consciously transferred data, when a timely and clear notification has not been provided, and therefore it cannot be assumed that a person is aware that data collection and storage is taking place.

One of the more active institutions in the field of big data ethics is the EDPS, which has addressed the difficulties around the use of big data and the respect of fundamental rights in various occasions. [19] Some of the main challenges envisaged by the EDPS when using big data have been described as:

1. *The lack of transparency*. This may be a problem where the PolicyCLOUD platform is used to create public policies, since the impacts on individuals directly derived from the adoption of new regulations will be based on the data gathered through the PolicyCLOUD platform. As noted in Section 2.1.4 below, the explainability of the functioning of the platform towards end-users (in particular, around how the platform generates specific outputs from the data collected) will be instrumental in ensuring that end-users can critically examine those outputs and make reasoned policy-making decisions which they can, in turn, explain to affected individuals. A key challenge of big data research is to justify and to develop value reference layers to big data. The usability of big data, for various purposes and targeted markets needs to be clarified.

2. *The informational imbalance between the holders of personal data and the corresponding data subjects*. Individuals may not be aware that their personal data is being processed through the PolicyCLOUD platform. This imbalance jeopardizes the ability of those individuals to exercise control over their

[7] E.g., data collected by the applications of mobile devices for which explicit consent was provided, data submitted during registration for the creation of a digital identity, etc.
[8] E.g., details and results of analyses collected by the first use case of the Project to realize counter-radicalisation policies.

personal data and enforce their rights under European law. The platform and its end-users must weigh the interests of the data subjects appropriately and find effective means to provide information about the activities performed on personal data, considering also the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the use case (e.g., the first use case of the Project).

3. *Unfairness and discrimination*. [20] Where the datasets on which the analysis relies on have been built with inherent prejudices or biases (e.g., where the data does not properly represent a target population, by overrepresenting one gender or ethnicity over another, or where the data draws correlations motivated by biases of the researcher), decisions made based on the analysis of those datasets may ultimately reflect those same prejudices or biases. Given the potential impact that policy-making decisions may have on individuals, as well as entire communities and societies, it is important to control data source quality, by ensuring that only reliable sources are used, and to routinely test the analytics components of the platform to ensure that they do not skew knowledge obtained from data in a biased manner.

One of the potentially most powerful uses of big data is to make predictions about what is likely to happen but has not yet happened and what we are likely to do but have not yet done. For example, big data in PolicyCLOUD would be used to predict a phenomenon of radicalization and the potential rate to commit crime (e.g., in the first use case). While the possibility to extract useful predictions may be an important advantage, it is important to bear in mind that an excessive reliance on massive amounts of data is at risk to becoming the "*pollution problem of the information age*", creating exposure to a "*dictatorship of data*" [21] where, according to one study by the Norwegian Data Protection Authority, "*we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicate our probable actions may be*". [21]

In the recent literature of big data research, an increasing section is dedicated to the capacity of big data to support social sciences research. There is the anticipation that big data is potentially a social good that must be secured and be used for the transparency of services, and for the evolution of a user-centric new culture for sustainable computing. In parallel, several concerns have been documented, mostly related to trust, privacy, personal data protection, and the protection of personality in the new technology-driven domain of services and applications. Furthermore, expected benefits of big data-based predictions may lead to overconfidence in its capabilities, reducing critical examination of the outputs of the platform by end-users. Big data applications may find spurious correlations in data, even in cases where there is no direct cause and effect between two phenomena that show a close correlation. In these cases, there is a risk of drawing inaccurate but also, when applied at individual level, potentially unfair and discriminatory conclusions.

The EESC [23] has identified several ethical problems deriving from the exploitation of big data that are common to ethical issues concerning AI. These problems are the following:

1. *Privacy and protection of personal data*. As noted in Section 2.2.4 below, the protection of privacy and personal data is fundamental towards a legal and ethical development of the platform, since there is a huge amount of data, including personal data, which may be stored and reused through the PolicyCLOUD platform.

2. *Tailored reality and the filter bubble*. Since the information gathered by the platform will be classified and clustered, knowledge extracted by the platform may paint only a partial picture of a context which may be more complex. The resulting vision of the problems might become progressively limited, even producing an echo chamber effect, with a progressive narrowing of the same vision. [24] [24]This situation has

relevance when reasoning about the policies to be adopted in a city[9] or in the case of policies related to unemployment.[10]

3. *After death data management*. This aspect relates to the use of data related to deceased individuals, and the governance of those data, including whether heirs to the individual should retain control over the deceased person's data, and how this control can be exercised.

4. *Algorithm bias*. As seen also in Section 2.1.4 below, interpretations supplied by algorithms are essential for obtaining useful information from big data, and it is therefore essential to be able to address any potential biases inherent to those algorithms which may compromise the accuracy and value of that information.

5. *Equal rights between data owner and data exploiter*. This speaks to the imbalance between the positions of the data subjects, i.e., the data owners, and those who collect and manipulate their data, i.e., the data exploiters. The implementation of effective mechanisms within the platform to allow individuals to exercise their rights around personal data, whether directly against PolicyCLOUD, or against platform end-users, is one core way through which this imbalance can be addressed, along with the provision of clear and intelligible information to end-users and to the public at large on how the platform may handle data, including personal data, so that it is possible for an average person to understand how the platform works, and how it may be leveraged by policy-makers.

From a different standpoint, the big data ecosystem requires distribution and aggregation of information in modes that were unforeseen in the past. The sophistication and the huge capacity of big data services to process significant volumes of data, automatically and without human intervention, sets critical questions related not only to privacy and personal data protection, but more specifically to security. The PolicyCLOUD platform should thus incorporate adequate technical and organisational security measures, developed as a result of a dedicated security risk assessment targeting potential threats generated, in particular, from reliance on big data analysis.[11]

It is also important to bear in mind, considering the data sources identified as relevant for the Project, the sheer amount of information diffusion taking place in the context of social networks and social media. The ease of sharing information as well as the increased openness of such data warehouses permits advanced data processing that leads to critical insights about the data providers (i.e., social media users). In this situation, big data applications serve as intermediaries, matching the gap between the providers and the consumers of data, allowing several innovative business models to appear. The power of big data applications built around new information processing methods such as sentiment analysis and opinion mining may allow for an extensive extraction of knowledge from social media user activity which may be unknown even to social media users themselves. If leveraged effectively, this knowledge could potentially be used to manipulate or coerce social media users into making decisions deemed as beneficial by the data exploiter, with an impact that could potentially affect the very functioning of democracy itself. [25] PolicyCLOUD must be aware of this when engaging with end-users for use of the platform and in particular when defining intended purposes for big data processing; on the other hand, end-users must ensure that the purposes for which they wish to leverage the functionalities of the platform are legally and ethically sound.

[9] See the third use case of the Project.
[10] See the fourth use case of the Project.
[11] For more on this, see Section 2.2.4 below.

Advanced user profiling is also critical for the launch and management of socially sensitive applications powered by big data research. The standardization of profiles is the first step toward interoperability of applications and social services. To this direction, latest developments in computer science as well as in policy-awareness frameworks provide significant contributions. From a social impact perspective, a key question is how, within governmental institutions and regulations, can we envision trustable, participatory, and democratic platforms that exploit big data profiles for social good. Social rating systems or social filtering platforms are key examples for this emerging area of research. Furthermore, from a social perspective, another key concern is about the ownership of the big data. Also, smart cities research is an example of critical integration for social sciences and computer sciences research. In all these cases several research questions link big data research to critical social impact.

Within this complex big data ecosystem, individuals, organizations as well as governments need to develop frameworks to measure their readiness for the integration of big data research for measurable individual and social objectives. One direction for the exploitation of big data research is analytics. The exploitation of value through huge volumes of data requires the development of big data analytics capabilities, aiming to provide visualizations and summaries of data that can promote enhanced decision making. From a social science perspective, this connection directly leads to a new era of smart urbanism, where human actors (e.g., citizens) exploit processed data in meaningful visual forms for the improvement of the quality of their lives.[12]

2.1.4 General ethical and societal considerations on AI

Potential solutions for ethical and societal issues related to AI are constantly discussed. The AI-HLEG is just one example of a working group focused on the need to create an appropriate AI liability framework. [26] In AI processing, ethical issues are vital, since alignment with ethical values is seen as one of the pillars to achieve a trustworthy AI system.[13]

The AI-HLEG Ethics Guidelines for Trustworthy AI (the "**AI-HLEG Guidelines**") set seven key requirements that AI systems should meet to be deemed trustworthy and which can be broken down into multiple considerations:

1. *Human agency and oversight*. AI systems should, ultimately, serve to guide end-users in their decision-making. As such, it is essential for output generated by AI-powered analysis of data to be understandable by end-users, in terms of results and the process followed by the AI system to reach those results. AI systems should not, conversely, be used to decrease, limit, or misguide human autonomy. Overreliance on the AI-generated outputs of the platform should be fought, by emphasising the responsibility of the end-users to critically examine those outputs and make reasoned decisions about how to incorporate them

[12] Another key aspect of big data literature is related to the big data hype. The use of big data research for social purposes must identify opportunities, myths as well as risks. It is necessary for our societies and for policy making purposes to ask appropriate questions related to the ownership, supervision, consumption, and protection of big data. Smart cities and smart government research must consider several delicate issues related to privacy, personal data protection, security, safety and social responsibility of individuals and groups. Without a focus on sustainability, social inclusive economic growth and social justice, any isolated, monolithic big data application in the long term will unfortunately fail to promote its social impact. Novel approaches are required in the management of big data and their interoperability, as well as the annotation of data and services for improved social services.

[13] According to the AI-HLEG Guidelines, trustworthy AI should be: 1. lawful, respecting all applicable laws and regulations; 2. ethical, respecting ethical principles and values; 3. robust, both from a technical perspective while taking into account its social environment.

into policies, noting that technically acceptable results may not always be socially or politically acceptable. AI systems should also not be leveraged to manipulate, deceive, herd or condition individuals, and thus PolicyCLOUD and end-users are responsible for ensuring the platform is not used for illegal or unethical purposes.

2. *Technical robustness and safety*. It must be assessed whether the AI system may create adversarial, critical or damaging effects (e.g., to human or societal safety) in case of risks or threats to its robustness, such as design or technical faults, defects, outages, attacks, misuse, and inappropriate or malicious use. A comprehensive security risk assessment must be performed, with relevant risks and threats identified and properly mitigated, as a continuous process (during the design and throughout the entire lifecycle of the PolicyCLOUD platform). Compliance with relevant cybersecurity certifications (e.g., in Europe, according to the certification scheme created by the CSA) [27] or other specific security standards may be instrumental in this respect.

3. *Privacy and data governance*. This represents a point of connection between ethical/societal and legal/regulatory requirements, and it is addressed further below, in Section 2.2.4.

4. *Transparency*. Traceability is a key ethical consideration for AI systems: all decisions made by an AI system should be logged, including the decision-making process followed (e.g., what data was used, what algorithm was applied, etc.). This is fundamental towards ensuring the explainability of AI outcomes to end-users and individuals affected by AI-influenced policies, without which AI outcomes cannot effectively be challenged. This must be supported by an adequate communication of the capabilities and limitations of AI systems to end-users, in terms of their intended purposes, the conditions under which they can function normally, and expected accuracy levels. Furthermore, end-users should ensure that they clearly communicate to affected individuals that a policy has been developed with the assistance of AI-based systems, explaining how their decisions were made.

5. *Diversity, non-discrimination, and fairness*. The potential for bias and discrimination in AI-based systems is magnified when compared to human decision-making, as AI systems are able to carry out decisions in a totally different scale in terms of quantity and speed, without being subjected to the same social controls as human decision-makers. Thus, it is fundamental to tackle bias and discrimination, in both datasets and algorithms used to analyse those data sets, from a design stage. This will include, during development, ensuring that sufficiently reliable and representative training data is used to develop AI systems, along with oversight mechanisms around algorithm development to control for bias and discrimination in output generation. After deployment, end-users remain responsible for ensuring the reliability of the data sources they may wish to leverage through the platform.

6. *Societal and environmental well-being*. This requirement implies the need to focus on the potential impacts of the PolicyCLOUD platform on the environment, such as the amount of energy used and related carbon emissions. For this reason, it is recommended to define measures to reduce the environmental impact of AI systems implemented via the platform throughout their lifecycle. Furthermore, end-users must bear in mind that reliance on the platform may have different societal implications depending on the context in which they intend to operate. Where the rights of citizens may be most directly affected, such as where the platform may be used to make decisions related to law enforcement, employment, social systems and structures or the democratic process, end-users must perform a comprehensive societal impact assessment to determine whether the risks of relying on an AI-based solution for guidance in policymaking are properly mitigated and are outweighed by the presumed benefits.

7. *Accountability*. Mechanisms facilitating the auditability of AI systems (e.g., traceability of the development process, sourcing of training data, and logging of the AI system processes, outcomes, and positive and

negative impacts) should be put in place. Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented. In this regard, leveraging the Assessment List for Trustworthy Artificial Intelligence defined by the AI-HLEG [28] to account for measures taken to address relevant ethical concerns, to the extent that it is applicable to the specific systems implemented within PolicyCLOUD, may be beneficial.

Adherence to these requirements may allow the AI systems implemented within the platform to be developed in a trustworthy manner, provided that PolicyCLOUD acknowledges that this effort requires a continuous identification of requirements and evaluation of solutions towards improved ethical outcomes throughout the AI system's lifecycle, with the collaboration of relevant stakeholders, such as the end-users. [28]

Other than the AI-HLEG, the EGE has called attention to the risks inherent to uncoordinated and unbalanced approaches in the regulation of AI and other autonomous technologies. In this direction, the EGE has proposed a set of basic principles and democratic prerequisites, based on the fundamental values laid down in the EU treaties and in the CFREU. [30]

More recently, the AI4People taskforce has surveyed the EGE principles as well as thirty-six other ethical principles put forward to date and subsumed them under four overarching principles. [31]

The EDPS has also launched a public consultation on digital ethics in 2018, the results of which highlighted many new challenges, such as the current inability of individuals "*to benefit from their data themselves*", the presence of biases in algorithms and the potential discrimination, the transparency and reproducibility of AI, the patents and copyright issues, the growth of fake news, online fraud, and cyber-bullying, which may impact on the "*goodness*" of the information gathered and the related adopted policies. [32]

As an essential part of the implementation of the abovementioned requirements to achieve a trustworthy AI, respect for and protection of the fundamental rights of individuals must be considered. Measures to safeguard fundamental rights during the development and deployment of the platform must be considered. To identify possible risks and appropriate mitigation, a FRIA should be performed. Based on the CFREU, on the ECHR and its protocols, and on the European Social Charter, the FRIA should be focused on issues such as:

1. Whether the AI systems may potentially be used to discriminate against people on the basis of any grounds, such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age, or sexual orientation.

2. Whether processes can be put in place to test and monitor the AI systems for discrimination during the development, deployment and use phases of those systems. Where actual or potential discrimination is detected, specific measures must be identified and put in place to address and rectify the data, the algorithms and/or any other components of the AI decision-making process, to ensure that the system produces outputs which are not discriminatory in a manner contrary to European values.

2.2 Legal and regulatory issues

In this section we will explore existing literature on the legal and regulatory issues related to the use of cloud systems, big data and data driven policy making, also considering the statements and guidelines issued on this topic by relevant European and international institutions.

The functionalities of the various foreseen PolicyCLOUD components presuppose the collection and further processing of data, which in turn requires the selection of appropriate sources from which those data may be collected.

From a practical perspective, whether a data source may be considered as appropriate depends on the purpose for which PolicyCLOUD is to be leveraged in relation to the data included in that data source. In other words, a data source will only be practically appropriate if it contains data which, after its processing through PolicyCLOUD, can be useful towards the goal which the PolicyCLOUD end-user wishes to achieve (i.e., the development of data-based policies for a specific use case).

However, the use of a data source which is practically appropriate may fall short of applicable ethical, legal, regulatory and/or societal requirements. A failure to comply with such requirements may, depending on the requirements in question, create various risks for PolicyCLOUD and the PolicyCLOUD end-user, as well as for other relevant stakeholders and individuals. Thus, data sources should not be selected solely based on their practical usefulness, but also on whether they can be lawfully and ethically leveraged for the goal which the PolicyCLOUD end-user wishes to achieve.

Particularly, as all use cases involve the processing of personal data, as defined by the GDPR, the main aspect which shall be considered in the development of each use case is the compliance with the applicable data protection legal and regulatory framework.

2.2.1 Contractual protection of data sources

Access to and use of certain data sources, particularly those hosted on online websites or platforms, may be governed by contractual terms. Depending on the applicable law, use of such data sources in a manner which is forbidden by their corresponding terms may be unlawful, potentially amounting to a breach of contract which may be enforceable against users through legal claims.

This is particularly relevant for data sources which can only be accessed upon active acceptance of such contractual terms. Assuming the terms themselves are compliant with the applicable law, those terms are likely to be considered legally binding towards users which accept them.

Where contractual terms governing use of a given data source exist, but the data source can be accessed without actively accepting them, whether those terms are legally binding towards the users is dependent on the applicable law. In particular, factors such as the manner in which those terms are presented to users during the use of the data source, the manner in which the user accesses the data source (e.g., whether the user accesses the data source normally, or merely accesses the data through use of a web crawler), and whether there are any technical measures in place to prevent certain uses of the data source (e.g., robots.txt file configured to prevent web crawling) will be considered.

If no contractual terms are applicable to a given data source, this requirement does not, in principle, apply to that data source.

Any data source which is selected for registration in PolicyCLOUD must, therefore, be assessed, prior to registration, to determine:

1. Whether any contractual terms governing access to and use of the data source exist.
2. If so, whether such terms prevent use of the data source through PolicyCLOUD, in the manner intended by the PolicyCLOUD user.
3. If so, whether such terms may be considered legally binding upon the PolicyCLOUD user.

Should a selected data source be subject to contractual terms which prevent its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the data source owner (i.e., the person or organisation entitled, under the contractual terms, to authorise the use of the data source).

2.2.2 Legal protection of databases

Regardless of contractual terms which may apply, data sources which can be classified as databases under Art. 1, par. 2 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 (the "**Database Directive**") may be the object of further legal restrictions on their use.

A data source will be classifiable as a database under the Database Directive if it is "*a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means*".[14] This is a rather broad definition which is likely to encapsulate most, if not all data sources which may be selected for registration on PolicyCLOUD.

Where the selection or arrangement of the contents of a database is carried out in a way which is original and creative, such that it can be considered the intellectual creation of its author and not just a methodical and systematic compilation of data, that database may be protected under copyright, under Art. 3 Database Directive. This may be true even if the contents of the database itself are not eligible for copyright protection (e.g., a database containing factual information). In this case, the holder of the copyright over the database may be entitled to restrict, among other acts, the reproduction, alteration, or communication, in whole or in part, of the database, under Art. 5 Database Directive.

Nevertheless, if the maker of a database has made a substantial investment in the obtaining, verification, or presentation of the contents of a database, that database may be protected under the database sui generis right, under Art. 7 Database Directive. In this case, the database maker may be entitled to restrict, among other acts, the extraction and/or re-utilization of the whole, or of a substantial part, of the contents of that database.

Exceptions to the above entitlements are also foreseen in the Database Directive, in Arts. 6 (relating to copyright) and 8 to 10 (relating to sui generis right).

The protection offered by the Database Directive is conditioned by the fact that, as a directive, it must be transposed into the national legislation of EU member States before becoming fully applicable in those member States, as opposed to a regulation. Since there is a level of discretion afforded to member States when transposing

[14] Art. 1, par. 2 Database Directive.

directives into their local laws, the extent of protection offered in each Member State may vary, insofar as it does not conflict with the terms of the Database Directive.

Any data source which is selected for registration in PolicyCLOUD must, therefore, be assessed, prior to registration, to determine:

1. Whether the data source may be considered as a database under the Database Directive.
2. If so, whether the database is eligible for protection under copyright or the sui generis right, under the Database Directive.
3. If so, whether the use of the database as intended by the PolicyCLOUD user falls under any exception foreseen in the law applicable to the database.

Should a selected data source be eligible for database copyright or sui generis right protection which prevents its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the rights holder (i.e., the person or organisation entitled, under the law applicable to the data source, to authorise the use of the data source).

2.2.3 Copyright

Copyright protection may extend to data sources which may not be qualifiable as databases under the Database Directive or to specific contents within those data sources.

As set out in the Berne Convention for the Protection of Literary and Artistic Works (the "**Berne Convention**"),[15] any data source or content which might be qualified as a "*literary* [or] *artistic work*"[16] may be protected, in the sense that the holders of copyrighted works automatically[17] enjoy exclusive rights over such works, including the

[15] WIPO, *Berne Convention for the Protection of Literary and Artistic Works*, https://wipolex.wipo.int/en/treaties/textdetails/12214, retrieved 2020-12-19.

[16] Under Art. 2, par. 1 Berne Convention, any "*production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression, such as books, pamphlets and other writings; lectures, addresses, sermons and other works of the same nature; dramatic or dramatic-musical works; choreographic works and entertainments in dumb show; musical compositions with or without words; cinematographic works to which are assimilated works expressed by a process analogous to cinematography; works of drawing, painting, architecture, sculpture, engraving and lithography; photographic works to which are assimilated works expressed by a process analogous to photography; works of applied art; illustrations, maps, plans, sketches and three-dimensional works relative to geography, topography, architecture or science*" may potentially qualify for copyright protection.

[17] Art. 5, par. 2 Berne Convention specifies, as a rule, that "[t]*he enjoyment and the exercise of* [exclusive rights under copyright] *shall not be subject to any formality*", implying that they arise automatically upon creation of a "*literary or artistic work*".

right of reproduction[18], broadcasting and public communication[19], adaptation and arrangement[20], among others. These rights are further developed and specified in several EU legal instruments, including Directive 2001/29/EC of the European Parliament and of the Council, of 22 May 2001 (the "**InfoSoc Directive**"),[21] but given the lack of an overarching EU Regulation on copyright, each Member State has its own set of rules on copyright protection, which, though assumedly compliant with the Berne Convention and applicable EU Directives, still present a significant degree of variance from one another.

In general, in the absence of an applicable exception to the exclusive rights afforded to the holder of copyright, any reproduction, communication, adaptation or arrangement of a copyrighted data source or part of a data source must be authorised by the holder of copyright. The following are examples of potentially relevant exceptions which may be foreseen under national legislation: [33]

1. Reproductions of works by libraries, archives, and museums.
2. Works used as illustration for teaching or scientific research.
3. Use of works for public security purposes.
4. Use for the purpose of research or private study.

Any data source selected for registration in PolicyCLOUD must, therefore, be assessed, prior to registration, to determine:

1. Whether the data source, or any relevant part of that data source, which is to be extracted, may be considered as protected by copyright, under EU and applicable member State law.
2. If so, whether the use of the data source as intended by the PolicyCLOUD user falls under any exception foreseen in the law applicable to the data source. This will naturally depend on the purpose for which the data source is to be used (i.e., the purpose for which the user wishes to rely on PolicyCLOUD).

Should a selected data source, or a relevant part of that data source, which is to be extracted, be eligible for copyright which prevents its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the rights holder (i.e., the person or organisation entitled, under the law applicable to the data source, to authorise the use of the data source).

[18] See Art. 9 Berne Convention. Art. 2 InfoSoc Directive defines the reproduction right as the exclusive right to authorise or prohibit direct or indirect, temporary, or permanent reproduction by any means and in any form, in whole or in part, of a copyrighted work.
[19] See Arts. 11-*bis* and 11-*ter* Berne Convention. The rights of communication to the public and of making available to the public are defined by Art. 3 InfoSoc Directive as the exclusive right to authorise or prohibit any communication to the public of copyrighted works, by wire or wireless means, including the making available to the public of copyrighted works in such a way that members of the public may access them from a place and at a time individually chosen by them.
[20] See Art. 12 Berne Convention.
[21] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

2.2.4 Personal data protection and privacy

This section aims at setting out the general legal principles underlying the processing of personal data in the context of the Project, while Section 3 highlights particularities arising from the application of these principles to the platform's different components, and Sections 4 to 7 below consider the peculiarities of each use case and therefore consider how each principle shall be concretely complied with.

2.2.4.1 LAWFULNESS PRINCIPLE

As determined by Art. 5, par. 1, let. a) GDPR, personal data must be processed lawfully. This means that any use of personal data must be performed on the basis of consent provided by the individuals whose data is used ("**Data Subjects**"), or otherwise on some other legitimate basis laid down in law, as set out in the GDPR or in other EU or member State laws referred to by the GDPR.

PolicyCLOUD must assess which of the legal bases afforded by the GDPR may be applicable and implementable for an intended processing of personal data. This assessment must consider the full context of the processing activities which are intended, including the specific data sources to be used, and the specific goals to be reached using the platform.

According to the GDPR, processing of personal data shall be considered as lawful only when:

1. Data Subjects have given their consent to the processing of their personal data for one or more specific purposes.
2. Processing is necessary for the performance of a contract to which the Data Subject is party or to take steps at the request of the Data Subject prior to entering a contract.
3. Processing is necessary for compliance with a legal obligation to which the Consortium is subject.
4. Processing is necessary to protect the vital interests of data subjects or of another natural person.
5. Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority.
6. Processing is necessary for the purposes of the legitimate interests pursued by the Consortium or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the Data Subject which require protection of personal data.

It follows that, for each processing activity carried out for the execution of the Project, a suitable legal basis for the processing must be identified.

Some general guidance can be provided on the particularities of specific legal bases which may potentially be relevant for consideration:

1. *Legal obligation.* One legal basis which may potentially be applicable is the need to process personal data to comply with a legal obligation to which end-users are subject.[22] This legal obligation should be laid

[22] Art. 6, par. 1, let. c) GDPR.

down in EU law or member State law applicable to the end-users, provided that the law in question meets an objective of public interest and is proportionate to the aim pursued.[23] As explained by the WP29 [34], the obligation in question must be imposed by law. The law must fulfil all relevant conditions to make this obligation valid and binding upon the controller, and must also comply with data protection law, including the requirement of necessity, proportionality, and purpose limitation. The legal obligation itself must be sufficiently clear as to the required processing of personal data, referring explicitly to the nature and object of the processing. Laws of non-EU countries are not covered by this legal basis. There must further be no choice on the controller but to comply with this legal obligation, nor should there be an undue degree of discretion afforded to the controller on how to comply with it. It must be necessary to process personal data to comply with a legal obligation upon the controller, for that legal obligation to potentially act as a legal basis for that processing. If it is possible to comply with the obligation without processing personal data or by processing fewer personal data than what is envisioned, this will not apply. It is recommended for end-users to assess whether this legal basis may apply to the processing activities it may envision performing through the platform, as a first step.

2. *Consent*. Where the above legal basis is unavailable, another which may apply is consent provided by Data Subjects for the use of their personal data, for the purpose of generating aggregated information from which the end-user may be able to draw insights into relevant trends and issues, so as to use those insights to focus policymaking on the most pressing issues and improve decision-making efficiency and effectiveness. To obtain valid consent from Data Subjects for this processing of their personal data, end-users should consider the feasibility of requesting consent from data subjects in a manner which meets all of the consent requirements of the GDPR, as further developed by the EDPB) [35]:

   - *Freedom of consent*. Consent will only be valid when it is freely given, i.e., where the Data Subject has a genuine choice in whether to provide his/her consent for the processing of their personal data for a given purpose. If a Data Subjects feel compelled to consent or are subject to negative consequences if they do not consent, then this consent will not be valid. If this consent would be made mandatory to allow data subjects to benefit from the provision of a service, this would affect the freedom of their consent and, therefore, its validity. If the Data Subjects are not allowed to refuse, or later withdraw, their consent, it will also not be considered freely given. Given the potential inherent imbalance of power between end-users (i.e., public entities) and individuals, particularly if those individuals are citizens under the jurisdiction of the end-user, it must be made absolutely clear to individuals that they will suffer no adverse consequences if they choose not to provide their consent, or to later withdraw it.

   - *Specificity of consent*. Consent will only be valid when it is specific, i.e., when the request for consent refers to a specific, explicit, and legitimate processing purpose. To meet this requirement, the consent request must be formulated in a manner which clearly indicates the reason for which consent is requested, and which allows an average Data Subject to understand exactly what he/she is agreeing to when providing his/her consent. A single request for consent should only refer to a single processing purpose: overly broad consent requests, which ask Data Subjects to agree to several purposes for which their personal data may be used, will be invalid due to lack of specificity.

[23] See Art. 6, par. 3 GDPR.

- *Transparency of consent.* Consent will only be valid when it is informed, i.e., when the Data Subject is provided enough information to be able to understand what he/she is consenting to, and what the implications of providing or refusing his/her consent may be. This requirement is tied to the principle of transparency. While the end-user remains obliged, under the principle of transparency, to provide Data Subjects with the full list of information required by Arts. 13 and 14 GDPR, it is possible to identify a minimum set of information which must be given to Data Subjects in connection with a consent request to ensure that it meets this requirement:

  a) The identity of the relevant controller, i.e., the end-user.

  b) The purpose for which consent is requested.

  c) The types of personal data which will be collected and used for that purpose.

  d) The existence of the right to withdraw consent, and how that right can be exercised.

  In specific cases, further minimum information may need to be provided.[24] This information should ideally be provided in writing, using clear and plain language, so that it is easily understandable for the average Data Subject. The consent request itself should be clear and distinguishable from other matters addressed to the Data Subject (i.e., it should be easy for the data subject to detect the consent request and decide on whether or not to provide their consent).[25]

- *Unambiguity of consent.* Consent will only be valid when it is unambiguous, i.e., where it reflects an indication of the agreement of the Data Subject to having his/her personal data processed for a specific purpose, provided by a statement or clear affirmative act from the data subject. This implies that the Data Subject must have taken a deliberate action to provide their consent, such as ticking a box or signing a document. Use of pre-ticked consent boxes, or reliance on implied consent through the silence, inactivity, or mere proceeding with a service of the Data Subject will not be considered valid means of obtaining consent under the GDPR. In particular, the action to provide consent must be distinct from any action made by the data subject to accept general terms and conditions for the use of a service.

- *Demonstrability of consent.* While not a requirement to ensure the validity of consent, the ability to demonstrate that consent has been provided is fundamental to ensure compliance with the principle of accountability. Adequate logs of consent provided must be kept by the end-user, showing:

  a) Who provided consent.

[24] In particular, if: 1. any automated individual decision-making activities, under Art. 22 GDPR, are to be performed using the personal data of the Data Subject, the existence of such activities, as well as meaningful information about the logic involved, the significance of those activities and the envisaged consequences of those activities for the Data Subject; and 2. if the personal data is to be transferred outside of the EEA, on the basis of the consent of the Data Subject, to a country that has not received an adequacy decision from the EC and without appropriate safeguards, under Art. 46 GDPR, having been implemented to govern the transfer, information on the possible risks which may arise for the Data Subject as a result.

[25] See Art. 7, par. 2 GDPR.

b) When consent was provided.

c) What was consented to.

d) How consent was provided.

e) What information was made available to the data subject prior to consent being given.

- *Explicitness of consent*. Where consent is relied on as a derogation to the general prohibition on the processing of special categories of personal data, it will only be valid when it is explicit, i.e., where the Data Subject has provided an express statement of consent, such as through a written statement, the filling in of an electronic form, the sending of an e-mail or the electronic signing of a consent statement.

  Consent will only be a feasible legal basis for the processing of personal data where all the above requirements can be met in practice. Furthermore, reliance on consent implies the need to ensure that consent given can be withdrawn, at which point all processing of personal data related to the consenting data subject must cease[26] and, in the absence of a legal basis to further process those data, deleted[27]: as such, consent can only be used where it is feasible to allow for the personal data of one or more Data Subjects to be deleted or removed from further processing, should their consent be withdrawn.

3. *Legitimate interests*. While this legal basis, available under Art. 6, par. 1, let. f) GDPR, is the most flexible out of the six legal bases available to controllers, it is mandatory for controllers to perform a specific assessment (referred to as a "balancing test" or a "legitimate interests assessment") to determine whether it can be leveraged. This will only be the case where the interests which a controller wishes to pursue with a given processing activity are not overridden by the interests or fundamental rights and freedoms of the data subjects concerned. However, it is key to note that Art. 6 GDPR does not allow public authorities, in the performance of their tasks, to rely on this legal basis. As such, if an end-user is acting as a public authority, in the performance of tasks mandated to it by law or regulation, as opposed to acting in the capacity of a private entity, this legal basis cannot be relied on by the end-user. Where Art. 6, par. 1, let. f) GDPR is available, as mentioned, a specific assessment must be performed. To put this in more practical terms, should end-users wish to leverage their own legitimate interests as a legal basis, end-users will be responsible for making sure that they are pursuing interests which are lawful, in a manner which does not excessively intrude upon the rights of Data Subjects. To accomplish this, end-users must carry out and document an assessment in which they balance their interests against those individuals' rights. Extensive guidance on the performance of these assessments has been given by the Article 29 Data Protection Working Party, in the context of the Data Protection Directive. [34] In general, the following practical steps are recommended in these assessments:

   - End-users should describe the intended activity, identifying relevant persons in charge of the activity and the systems used in connection with the activity. It should be clarified whether the intended

[26] See Art. 7, par. 3 GDPR.
[27] See Art. 17, par. 1, let. b) GDPR.

activity will require the processing of personal data and, if so, the specific categories of personal data should be identified.

- End-users should establish whether Art. 6, par. 1, let. f) GDPR is the most appropriate legal basis for the activity in question.
- End-users should describe the interest being pursued. As noted by the WP29 "[t]he concept of 'interest' is closely related to, but distinct from, the concept of 'purpose'", [34] whereas a "purpose" is the specific reason for which personal data are processed, and an "interest" is the broader stake that the controller may have in the processing activity, or the benefit which may be derived from this activity. [34]
- End-users should establish whether this interest is lawful, in that it does not amount to the pursuit of illegal values or goals, and whether it is a real and present interest of the end-user, as opposed to overly vague or speculative interests. [34]
- End-users should assess the specific purposes for which personal data will be processed. Each purpose must be described, and it must be determined whether the intended processing activity is strictly necessary to meet the purpose. In essence, this requires end-users to make an impartial and comprehensive assessment as to whether there is any less-intrusive way its goals could be achieved.
- Interests pursued by end-users must then be assessed more in-depth: it is important to explain whether those interests correspond to the exercise of a fundamental right of the end-users or a third party, under EU law, whether they line up with the public interest or wider interests of the communities in which end-users are inserted, and whether they are legally, socially, and/or culturally recognised as legitimate. The impact upon end-users or third parties if the activity is not carried out is also relevant at this stage.
- End-users must consider the impact on the data subjects affected by the processing. Accordingly, it must be understood, for example:
- Whether any sensitive data28 are handled in connection with the activity.

At the end of this exercise, end-users should be able to provisionally conclude as to whether their interests manifestly outweigh the impact upon Data Subjects, or whether the impact upon Data Subjects is clearly excessive and disproportionate towards the aims sought by end-users, particularly where there may exist less intrusive alternatives to meet the same goal. Where this is not the case (i.e., no clear resolution either way is achieved), end-users must identify additional safeguards for the intended processing activity which aim to resolve the conflict in favour of end-users, by further ensuring that the rights, freedoms of interests of data subjects are adequately protected. [34]

[28] The concept of "sensitive data" used here is broader than that of "special categories of personal data", as established by Art. 9 GDPR. It includes those data, as well as personal data on criminal convictions and offences disciplined by Art. 10 GDPR, communications data, location data, financial data, and, in general, any information on individuals that may require special protection.

In any case, the safeguards put in place should sufficiently address the risks detected for the rights of the Data Subjects concerned, so that end-users may convincingly demonstrate that the interests they wish to pursue are not overridden by those rights. Only where this is possible it will be feasible for end-users to rely on their own legitimate interests or those of third parties as a valid legal basis under the GDPR.

4. *Public interest.* Where it is not feasible for end-users to implement a consent request mechanism in line with all the above requirements, and where the option of Art. 6, par. 1, let. f) GDPR is not available, end-users may consider whether they can justify the intended processing of personal data on the need to perform a task carried out in the public interest, or in the exercise of official authority.[29] Reliance on this legal basis presupposes the existence of a legal provision, though not necessarily a legal obligation, under EU law, or under Member State law applicable to end-users, which serves as a mandate for that task to be carried out, or that authority to be exercised, provided that the law in question meets an objective of public interest, and is proportionate to the aim pursued.[30] Tasks carried out in the public interest of a third country or in the exercise of official authority vested by virtue of foreign law do not fall within the scope of this legal basis. [34] This legal basis presents the most flexible approach available to public authorities, when acting in the performance of their tasks. Where the other legal bases mentioned are not available, end-users should assess whether the requirements for this specific legal basis are met, to ensure the lawfulness of the intended processing activities.

2.2.4.2 LAWFULNESS PRINCIPLE (SPECIAL CATEGORIES OF PERSONAL DATA)

Where special categories of personal data[31] are to be collected and further processed, an applicable derogation to the general prohibition on the processing of these personal data[32], from those listed in Art. 9, par. 2 GDPR, or as may be further provided under applicable member State law, must also be identified. For clarity, to lawfully process special categories of personal data, a controller must identify an applicable legal basis under Art. 6 GDPR and an applicable derogation under Art. 9 GDPR.

Some general guidance can be provided on the particularities of specific derogations which may potentially be relevant for consideration:

1. *Explicit consent.* Consent can serve as a legal basis for processing of personal data and as a derogation to the general prohibition on processing special categories of personal data, provided that it is explicit, i.e., where the Data Subject has provided an express statement of consent, such as through a written statement, the filling in of an electronic form, the sending of an e-mail or the electronic signing of a consent statement. On this point, we refer to the requirements set out in the previous section, as all of those must be met to ensure validity of consent.

[29] See Art. 6, par. 1, let. e) GDPR.
[30] See Art. 6, par. 3 GDPR.
[31] Art. 9 GDPR defines the special categories of personal data as "[…] *personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and* […] *genetic data, biometric data* [when processed for the purpose of uniquely identifying a natural person], *data concerning health or data concerning a natural person's sex life or sexual orientation*".
[32] See Art. 9, par. 1 GDPR.

2. *Data manifestly made public by the Data Subject*. Where a Data Subject has manifestly made personal data public, this may serve as a derogation to the general prohibition on use of special categories of personal data related to him/her.[33] However, there is a high threshold to be met for this derogation to be applicable.

   In the context of social media platforms, the EDPB has indicated the following factors to be considered in determining whether personal data has manifestly been made public [36]:

   - The default settings of the social media platform (i.e., whether the Data Subject took a specific action to change default private settings into public or not).
   - The nature of the social media platform (i.e., whether this platform is intrinsically linked with the idea of connecting with close acquaintances of the Data Subject or creating intimate relations, or if it is meant to provide a wider scope of interpersonal relations).
   - The accessibility of the page where the sensitive data is published (i.e., whether the information is publicly accessible or if, for instance, the creation of an account is necessary before accessing the information).
   - The visibility of the information where the Data Subject is informed of the public nature of the information that they publish.
   - If the Data Subject has published the sensitive data himself/herself, or whether instead the data has been published by a third party or inferred: where the Data Subjects provided the data themselves, and the collected data is directly extracted from the information provided by the Data Subjects, as opposed to being inferred from that information, this weighs in favour of this derogation.

   End-users should assess whether this derogation may be applicable to any of the data sources they wish to rely on, should the processing of special categories of personal data be intended.

3. *Substantial public interest*. This derogation requires a basis in EU or member State law applicable to the controller which must:

   - Be proportionate to the interest pursued.
   - Respect the essence of the right to data protection.
   - Provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests.

   The processing of special categories of personal data must be demonstrably necessary to meet that interest. [37] In general, the requirements analysed in the prior section for the legal basis of public interest must be met, added by these specifications to ensure that the interest pursued is, effectively, substantial.

   End-users should assess whether this derogation may be applicable, considering the purposes which they may be seeking to pursue through processing of personal data via the platform.

[33] See Art. 9, par. 2, let. e) GDPR.

4. *Statistical purposes*. Where special categories of personal data are processed for statistical purposes, based on EU or Member State law applicable to the controller which must be proportionate to the interest pursued, respect the essence of the right to data protection, and provide for suitable and specific measures to safeguard fundamental rights and interests of the Data Subjects, this derogation may be applicable. These specific safeguards are further specified in Art. 89, par. 1 GDPR, which imposes an obligation to implement technical and organisational measures to ensure respect for the principle of data minimisation, including pseudonymisation[34] of data or, whenever possible, the aggregation or anonymisation of data. This derogation is potentially the most flexible of those available to end-users and PolicyCLOUD, should the processing of special categories of personal data be envisioned. However, for it to be applicable, controllers must emphasise respect for the principle of data minimisation, seeking to minimise the amount of personal data collected, to anonymise and aggregate those data whenever possible and as soon as possible and to otherwise pseudonymise those personal data.

2.2.4.3 FAIRNESS PRINCIPLE

Processing of personal data shall always be fair, meaning that PolicyCLOUD should only handle personal data in ways that may be reasonably expected and not use such data in a way that may produce unjustified adverse effects on data subjects. This is particularly important with reference to the analysis tools and algorithms used by PolicyCLOUD to extract information from collected data.

The assessment on the fairness of processing arguably depends on how personal data has been obtained by PolicyCLOUD. This means that, should Data Subjects be deceived about how personal data will be processed, the reasons why they will be processed or, in general, should a processing activity carried out by PolicyCLOUD affect the interests of individuals, such processing activity will be likely considered as unfair.

However, personal data may sometimes be used in a way that negatively affects an individual without this necessarily being unfair. What matters is whether such detriment is justified. This is particularly important where the legal basis PolicyCLOUD relies on for processing is the legitimate interest, as the legitimate interest assessment which shall be performed to ensure that such legal ground may be used for a specific processing activity will also take into account this aspect.

The principle of fairness is also closely related to the data protection rights granted to data subjects by the GDPR. Indeed, pursuant to such principle, when Data Subjects seek to exercise their rights granted by the GDPR[35] or other applicable data protection laws PolicyCLOUD shall be capable to facilitate the exercise of these rights.

There are thus two core concerns surrounding the principle of fairness, enshrined in Art. 5, par. 1, let. a) GDPR:

1. The need to respect the reasonable expectations of Data Subjects.
2. The need to ensure that effective mechanisms exist to allow Data Subjects to exercise their rights.

[34] Pseudonymised data are still personal data, as they still refer to identifiable individuals, since pseudonymisation can be reversed.

[35] The rights of the Data Subjects listed by the GDPR are the right to information, the right of access, the right to rectification, the right to erasure, the right to restriction of processing, the right to data portability, the right to object and the right to not be subject to automated decision-making.

Personal data should not be processed in a manner that is detrimental, discriminatory, unexpected, or misleading for Data Subjects.

As noted by the EDPB [38][38], the following key elements illustrate those which should be considered to ensure that the principle of fairness is properly considered in the design of any intended activities around the processing of personal data:

1. *Autonomy*. Data Subjects shall be granted the highest degree of autonomy possible with respect to control over their personal data. This has an element of transparency, in that it implies that Data Subjects must be clearly informed as to what personal data of theirs may be processed, for what purposes it may be processed, in what manner it may be processed, what rights they have and how they can exercise them, but also a specific element of fairness in that Data Subjects must be provided means with which to effectively control what happens to their personal data, to the greatest extent feasible, including, necessarily, means by which they can exercise their rights under the GDPR.

2. *Interaction*. Data subjects must be able to communicate and exercise their rights with end-users.

3. *Expectation*. Processing should correspond with the expectations of the Data Subjects. If the Data Subjects are led to believe that personal data collected on them will be used to improve the policy-making abilities of the end-users, this should be the only objective pursued with those personal data, using them to profile and target individuals, or for other unrelated and arguably illegitimate purposes (e.g., sending of marketing communications), must be strictly avoided. This will imply controls around purpose limitation, including access control.

4. *Non-discrimination*. End-users shall not discriminate against Data Subjects. In particular, personal data should not be collected on Data Subjects for the purpose of discriminating against them, such as to cause harm or detriment to specific data subjects, nor should this be the end-result of policies developed using personal data. This requirement is strongly tied to applicable ethical considerations of avoidance of biases and non-discrimination.

5. *Non-exploitation*. End-users shall not exploit the needs or vulnerabilities of Data Subjects. Considering the potential imbalance of power between end-users (a public entity) and individual Data Subjects, this is particularly relevant when assessing the freedom of consent, where this legal basis is leveraged, in that data subjects should not be coerced or conditioned into providing their consent for use of their personal data for the purposes intended by the end-user under penalty of relevant detriment.

6. *Power balance*. Asymmetric power balances shall be avoided or mitigated when possible. This ties into the previous point: where consent is relied on, it must be made clear to Data Subjects that they will not suffer any negative consequences should they refuse to provide their consent, or later choose to withdraw it. Even where consent is not relied on, end-users must ensure that they comply with all applicable legal obligations when handling personal data and must develop policies based on those data with a reasoned and critical approach, having data subjects' fundamental rights and freedoms at the forefront of the decision-making process, to avoid abuse of power or arbitrariness.

7. *Respect for rights and freedoms*. End-users must respect the fundamental rights and freedoms of Data Subjects and implement appropriate measures and safeguards to not violate these rights and freedoms.

8. *Ethical processing of personal data*. End-users should see the wider impact of the processing on rights and dignity of data subjects. This creates, in effect, a legal obligation for end-users to bear relevant ethical and societal principles in mind, whenever personal data are handled, as seen in Section 2.1, above.

9. *Truthfulness*. End-users must act as they declare to do, provide account for what they do and not mislead the Data Subjects. Truthfulness is an essential requirement of transparency, which is not only a legal obligation regarding the processing of personal data, but also a prerequisite for the establishment of trust with Data Subjects and their acceptance of

    - The processing of their personal data for the purposes intended by end-users.
    - The subsequent policies which may be developed based on their personal data.

End-users must carefully assess the manner and extent to which they intend to collect and further leverage personal data through the platform, identifying specific measures to address each of the above points, to ensure that the principle of fairness is considered from the outset, by-design.

2.2.4.4 TRANSPARENCY PRINCIPLE

The transparency principle is strictly intertwined with the fairness principle, as it requires PolicyCLOUD and the end-users to be clear and honest with Data Subjects about the identity of the data controller which is collecting, processing and storing personal data, the methods used to process personal data, and the purposes of processing.

The obligation to comply with this principle applies regardless of whether personal data have been directly collected from data subjects or there is no direct relationship with individuals and collection of personal data comes from a third-party source. To some extent, where personal data comes from a source other than the Data Subject personal data refer to, transparency is even more important for PolicyCLOUD and the end-users, as data subjects may have no idea of the processing activities carried out by PolicyCLOUD and the end-users, whenever each party acts in their capacity as controller. By being open and transparent, PolicyCLOUD and end-users can show accountability towards the relevant Data Subjects and the community at large, publicly stating the terms under which they will process personal data. Naturally, in this manner, they may be held accountable by statements made.

Information should be provided efficiently and succinctly, to avoid information fatigue on the part of Data Subjects. It should be clearly differentiated from non-data protection related information. The language used should be duly considered to ensure that it can be understood by an average Data Subject, avoiding unnecessary ambiguities and describing the information in as simple a manner as possible, without resorting to complex sentence and language structures.

Ideally, this would involve the development of an information notice, to be provided directly to Data Subjects upon collection of their personal data, in writing. Such a notice must be developed with the inherent tension between the GDPR requirements of providing comprehensive information, and ensuring that the information provided is concise, transparent, intelligible, and easily accessible in mind. This requires an assessment as to which information should be prioritised, what the appropriate level of detail is, and which are the best means by which to convey this information to Data Subjects. [39] Whenever feasible, the so-called “layered approach” should be used, allowing information to be structured into relevant categories which the Data Subject can select, to ensure immediate access to the information deemed most relevant by the data subject and prevent information fatigue. [39] Where information is collected outside of an online context, one way to follow this approach would be to provide Data Subjects with an abbreviated paper-based notice at a data collection point, including a link to the more complete privacy statement made available online. [39]

Under the principle of purpose limitation, controllers must limit themselves to using personal data collected for the specific purposes identified at the time of collection, of for additional compatible purposes. [39] Whenever personal data can lawfully be processed for a further purpose (i.e., where this further purpose is compatible with

the original, or where an additional legal basis exists for the further purpose), the Data Subject must be given all relevant information as to that purpose, under Arts. 13, par. 3 and 14, par. 4 GDPR.

Any material or substantive changes to information notices, reflecting changes to the underlying processing activities, should be communicated directly to Data Subjects in a manner which ensures that they will be noticed, whenever feasible. [39] It will generally not be valid to merely inform Data Subjects that they should regularly contact a specific end-user or check an online information notice for changes or updates, given the inherent unfairness to Data Subjects which this represents. [39]

Where personal data is not collected directly from Data Subjects, and instead is collected from a third-party data source, there are circumstances under which an exemption from the obligation to provide information to Data Subjects directly may apply. In particular, this is not required where this proves impossible [39], or where the provision of this information would represent a disproportionate effort, particularly where personal data are processed for archiving purposes in the public interest, scientific and/or historical research purposes or statistical purposes, due to factors which are directly connected to the fact that personal data was not obtained directly from the data subject. [39] It is also possible to avoid this obligation where the provision of information would be likely to render impossible or seriously impair the achievement of the objectives sought by the processing activity. [39] In all of these cases, appropriate measures must be implemented to ensure the protection of the rights and freedoms of individuals regardless of the fact that this information is not directly provided to them, such as by displaying the information on a publicly available website, as stated in Art. 14, par. 5, let. b) GDPR.

One further exemption from this requirement applies where personal data is collected indirectly[36], if the collection of those personal data is expressly laid down in applicable EU or member State law, as laid down in Art. 14, par. 5, let. c) GDPR. This law must address the controller (PolicyCLOUD and/or a specific end-user) directly, making the collection of personal data mandatory: as such, the controller must be able to demonstrate how the law in question applies to them and requires them to collect the personal data in question. This collection should, in any case, be disclosed to Data Subjects, and it should be made clear that it is carried out in accordance with the law in question, unless there is a legal prohibition preventing the controller from doing so. [39]

As such, both PolicyCLOUD and the end-users, in relation to the processing activities which they may respectively perform, as controllers, shall therefore lay down a specific and easily accessible document which duly informs Data Subjects of the processing activities carried out in the context of the Project: a privacy policy. Such privacy policy, pursuant to the GDPR, shall inform individuals in a clear and plain language at least about the following:

1. The identity and the contact details of the controller.
2. The contact details of the DPO, where existing.
3. The purposes of processing for which the personal data are intended as well as the legal basis.
4. Where the data controller relies on the legitimate interest as the legal ground for processing, mention to such legitimate interest.
5. The recipients or categories of recipients of the personal data, if any.

[36] This exception does not apply to personal data collected directly from data subjects – Art. 13 GDPR applies fully to such cases.

6. Where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the EC about the level of data protection ensured by such third country or international organisation or, where the data transfer is based on other legal grounds pursuant to the GDPR[37], reference to such legal grounds.

7. The period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period.

8. The existence of the right for Data Subjects to request to the data controller access to and rectification or erasure of personal data or restriction of processing, or to object to processing as well as the right to data portability.

9. Where the processing is based on consent, the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal.

10. The right to lodge a complaint with a supervisory authority.

11. Specification as to whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the Data Subject is obliged to provide the personal data and of the possible consequences of failure to provide such data.

12. The existence of automated decision-making, including profiling, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the Data Subject.

In the case of PolicyCLOUD, this privacy policy should be made available on its cloud-based platform (see Section 3.3, below), with appropriate steps taken to make it available to the Data Subjects whose personal data are used in the context of the Project. End-users should likewise ensure that the above information is available to Data Subjects on public websites under their control, where direct provision of the information is exempted under any of the circumstances provided for in Art. 14, par. 5 GDPR, as explained above.

2.2.4.5 PURPOSE LIMITATION PRINCIPLE

Controllers shall collect personal data only for specified, explicit and legitimate purposes and not further process such data in a manner that is incompatible with those purposes. Such obligation aims to ensure clearness and openness about the purpose of processing, as well as to guarantee that the processing activities are in line with the reasonable expectations of Data Subjects.

Following this principle, platform users shall be limited both from a technical and from a contractual point of view in how they can process personal data which are collected and managed through the PolicyCLOUD platform. This also implies the need to implement internal policies within PolicyCLOUD to make users aware of what they can and cannot do with the personal data collected for the different use cases and more in general for the execution of the Project.

While the purpose limitation requirement applies for all processing activities, this may be considered as a paramount principle for the development of the Data Marketplace, as this will allow different datasets to be

[37] See Arts. 46 and 47 GDPR.

uploaded and subsequently used between PolicyCLOUD users, with the potential risk of using such datasets for purposes other than those permitted.

However, it should be noted that if the purpose limitation principle is to be considered as the general rule pursuant to the GDPR, this does not fully prejudice the possibility that personal data initially collected for one purpose may subsequently be used for different and additional purposes.

Indeed, on the one hand such further processing may be carried out if one of the following applies:

1. The new purpose is considered as compatible with the original purpose after having performed a specific compatibility test pursuant to the GDPR.[38]

2. Data Subjects have given their specific consent to the new purpose.

3. A legal provision exists which allows the new processing in the public interest.

One key aspect of this principle is the presumption established for "*further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes*" in Art. 5, par. 1, let. b) GDPR, such that any such further processing will be presumed to be compatible with the original purpose for data collection, provided that the safeguards of Art. 89, par. 1 GDPR are respected. While this presumption should not be seen as a general authorisation to further process personal data in all cases for historical, statistical or scientific purposes, as each case must be considered on its own merits and circumstances, it is in principle lawful for the further use of personal data collected in a commercial or healthcare context and arguably also in the context of the municipality contact centre to be further used for scientific research purposes and arguably also for statistical analysis aimed at the pursuit of a public interest, provided that the appropriate safeguards of Art. 89 GDPR are in place. [40]

The key design and default elements to be considered, for compliance with this principle, include [38]:

1. *Predetermination*. The legitimate purposes must be determined before the design of the processing. End-users must have established the specific purposes they intend to achieve before the mechanisms by which personal data are to be collected and used are designed and implemented.

2. *Specificity*. The purposes must be specific to the processing. It should be explicitly clear to data subjects why personal data is being processed.

[38] According to Art. 6, par. 4 GDPR, "*Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; (b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller; (d) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10; (e) the possible consequences of the intended further processing for data subjects; (f) the existence of appropriate safeguards, which may include encryption or pseudonymisation.*".

3. *Purpose orientation*. The purpose of processing should guide the design of the processing and set processing boundaries. This is particularly important in that the purpose will provide the baseline according to which all other processing principles may be complied with.

4. *Necessity*. The purpose determines what personal data is necessary for the processing. This ties into the principle of data minimisation. Where no personal data is necessary (i.e., where only aggregated or anonymous data would suffice), none should be used.

5. *Compatibility*. Any new purpose must be compatible with the original purpose for which the data was collected and guide relevant changes in design. The presumption of compatibility mentioned above applies here, for research and statistical purposes.

6. *Limitations to further processing*. Controllers should not connect datasets or perform any further processing for new incompatible purposes. This should be borne in mind by end-users when defining their intended purposes for use of the platform: they should only rely on data sources for which the purposes of the end-users have been clearly stated, which implies that connections between data sources involving personal data should not be carried out unless this has been made clear to Data Subjects, and an applicable legal basis has been found for these connections.

7. *Review*. End-user must regularly review whether the processing is necessary for the purposes for which the data was collected and test the design of the processing activities and the platform on which they are run against purpose limitation.

8. *Technical limitations on reuse*. PolicyCLOUD should implement technical measures, including hashing and cryptography, to limit the possibility of repurposing personal data.

End-users, in collaboration with PolicyCLOUD, should carefully assess the technical and organisational measures which can feasibly be implemented to address all the above requirements, to ensure compliance with the principle of purpose limitation.

2.2.4.6 DATA MINIMISATION PRINCIPLE

PolicyCLOUD shall only collect personal data, which is adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. As such, for each purpose of processing connected to the Project, it shall identify the minimum amount of personal data needed to fulfil such purpose. Compliance with the principle of data minimisation therefore requires a minimalistic approach to personal data, in the sense that

1. As little of it as possible should be processed to meet an intended purpose.

2. Only personal data which are adequate, relevant, and strictly necessary to meet a purpose should be used.

Ultimately, if a purpose can be met without using personal data, then no personal data should be used at all.

This requires end-users to consider, in the first place, whether it is possible to derive sufficient value from the data sources they wish to leverage through the platform without needing to preserve links between data and the individuals providing them.

In any case where end-users are not fully satisfied that retaining personal data is adequate, relevant and necessary to the purposes they wish to pursue, because it is still possible to pursue those purposes using anonymous or aggregated data, even if this is at the cost of efficiency, no personal data should be retained. Otherwise, end-users must be able to demonstrate the need for each data point collected on an identifiable data subject, considering the legal basis identified and the purposes pursued.

Data minimisation extends to each configuration of software and information systems within PolicyCLOUD, as from the design stage, used to process personal data in such a way as to reduce their use to a minimum (so-called "data protection by design"). Also, it applies to the development of technologies and/or processes with the aim of collecting and processing only the personal data strictly necessary to meet the purposes carried out through the development of the cloud-based platform by PolicyCLOUD, thus ensuring legitimate processing by default (so-called "data protection by default").

The key design and default elements to be considered, for compliance with this principle, include [38]:

1. *Data avoidance*. Avoid processing personal data altogether when this is possible for the relevant purpose. End-users must carefully consider the extent to which this may be feasible and rely solely on anonymous or aggregated data when it is.

2. *Relevance*. Personal data shall be relevant to the processing in question, and controllers shall be able to demonstrate this relevance. If end-users determine that the use of personal data, preserving a link to the identity of individual data subjects, is necessary, then end-users must be able to demonstrate that each data point collected is specifically relevant to the purpose pursued. Any irrelevant personal data will be deemed as excessive and should not be collected or further processed.

3. *Necessity*. Each personal data element shall be necessary for the specified purposes and should only be processed if it is not possible to fulfil the purpose by other means. This ties into the data avoidance element, at a smaller scale: the end-user should only use the strict minimum amount of data points needed to meet its purpose. If a specific data point is not strictly necessary ("nice-to-have", instead of "need-to-have"), it should not be collected or further processed, even if it could potentially be useful in the future.

4. *Limitation*. Limit the amount of personal data collected to what is necessary for the purpose. This ties into the necessity element.

5. *Aggregation*. Use aggregated data when possible. Aggregated data, where personal data from several different Data Subjects are combined to produce an output which no longer allows the linking of those data to any given individual, are not qualified as personal data under the GDPR, due to lack of identifiability of the underlying Data Subjects. As such, one way in which the principle of data minimisation can be respected, other than simply not collecting any personal data, is to promptly aggregate those data after collection.

6. *Pseudonymisation*. Pseudonymise personal data as soon as it is no longer necessary to have directly identifiable personal data, and store identification keys separately. Pseudonymised data is still personal data, as it is still possible to link those data back to the individual data subject through the identification key. However, to those without the identification key, pseudonymised data will, in principle, be unidentifiable, assuming the pseudonymisation is done appropriately, and thus relying on pseudonymisation reduces the likelihood of availability of personal data to unauthorised individuals, making it a useful measure also to ensure the principle of purpose limitation.

7. *Anonymisation and deletion*. Where personal data is not, or no longer necessary for the purpose, personal data shall be anonymised or deleted. Just as in the aggregation element, personal data lose their qualification as such when they are irreversibly and properly anonymised. [18] Whenever it is possible for end-users to meet their purposes without using personal data, reliance on anonymous data is an acceptable alternative. If, at any point, initially collected personal data are no longer necessary for a given purpose, they must be promptly aggregated, anonymised, or deleted.

8. *Data flow*. The data flow shall be made efficient enough to not create more copies, or entry points for data collection than necessary. From a technical perspective, this seeks to avoid unnecessary redundancy of collected personal data, to minimise the risk of personal data breaches. This element must be balanced against measures put in place to ensure data availability, such as back-ups.

9. *State of the art*. Controllers should apply available and suitable technologies for data avoidance and minimisation.

End-users, in collaboration with PolicyCLOUD, should carefully assess the technical and organisational measures which can feasibly be implemented to address all the above requirements, to ensure compliance with the principle of data minimisation. In any case, that the core initial assessment to be performed is as to whether any personal data at all needs to be used, or if end-users can feasibly and effectively meet their goals with only anonymous or aggregated data. The latter option is preferred from the legal perspective, and from a practical standpoint, as it reduces the compliance burden, by shifting away from the scope of privacy and data protection law.

2.2.4.7 ACCURACY PRINCIPLE

Controllers shall ensure the accuracy and quality of personal data collected, processed, and stored, being always sure to process up-to-date data. Ensuring accuracy of data used is fundamental from the legal perspective, as required by Art. 5, par. 1, let. d) GDPR, but also from the ethical and societal perspective, given that inaccurate, incomplete, misleading or biased data can result in erroneous outputs, culminating in misguided policymaking with a potential impact at an individual and societal level – as seen in Sections 2.1.3 and 2.1.4 above.

While this applies generally to all data which may be used through the platform, it has particular ramifications for personal data, under the GDPR's principle of accuracy. End-users and PolicyCLOUD will be required to ensure that appropriate steps are taken to verify the accuracy of any personal data collected, to maintain those personal data up to date over time, and to allow data subjects to correct, complete or update their own personal data when needed. This principle thus has:

1. An active component (measures which the controller must actively take to ensure data accuracy).

2. A passive component (an obligation for the controller to facilitate steps taken by data subjects to ensure data accuracy).

Such principle is particularly important where personal data are not directly collected from data subjects but from third-party data sources, as there will be a need to ensure that the source and status of personal data is clear. Ensuring accuracy when personal data used comes from a third party may be tricky, as it does not originate directly from the Data Subject. In these cases, it is important to accurately record each source of information, as well as challenge information gathered, if appropriate, trying to compare information collected by a third-party source with information available from another third-party source when reasonable doubts exist on whether the information collected is accurate and/or updated.

Accuracy is of fundamental importance for the success of the Project, since in case of outdated or incorrect information, the platform may not be able to generate valid output for any of the identified use cases, leading, in the best-case scenario, to partial or incorrect evaluations and, in the worst-case scenario, to consequences that could have adverse legal or other similar effects on the data subjects.

PolicyCLOUD shall therefore assess the analysis tools and algorithms used to ensure that information extracted is aligned with reality and not biased.

The accuracy principle is strictly connected with the right of rectification granted by the GDPR to Data Subjects. Pursuant to this principle, data subjects have the right to rectify their data and where their data are inaccurate or not updated, they have the right to obtain, as a precautionary measure, the limitation of processing for the entire period necessary for PolicyCLOUD to carry out the appropriate checks and rectification procedures where necessary. Finally, if it is not possible to update or rectify the data, Data Subjects have the right to obtain the erasure of such data.

PolicyCLOUD is therefore responsible for implementing technical and organisational measures aimed at guaranteeing the accuracy and quality of personal data included in the cloud-based platform, as well as for providing means to Data Subjects for contributing to the maintenance of data that is always accurate and up to date.

The key design and default elements to be considered, for compliance with this principle, include [38]:

1. *Data source reliability*. Data sources should be reliable in terms of data accuracy. As noted in Section 2.1.3 and 2.1.4 above, this is also an ethical requirement, in that use of erroneous or biased data sources may taint the policy-making process, with potentially substantial negative impacts on individuals and society at large. Though end-users may be assisted by the components of the platform in this respect, it is end-users which must take primary responsibility in ensuring that the data sources they select for registration on the platform are reliable and accurate. Only credible and trustworthy data sources should be leveraged, to mitigate the risk of poor-quality data being collected.

2. *Degree of accuracy*. Each personal data element shall be as accurate as necessary for the specified purposes. Essentially, this means that it is not required to ensure absolute or error-proof accuracy for all data collected, but rather that the likelihood or margin of error is acceptable for each data point, based on the purpose for which data is being processed and the potential impact which errors may have on individuals and society.

3. *Measurable accuracy*. Reduce the number of false positives/negatives. Measures should be implemented to mitigate the risk of errors not only in data collection, but also in outputs generated from further processing of those data. This will require a concerted effort from end-users and PolicyCLOUD: the platform will need to be able to explain the activities it has performed to collect and process data, and thereby generate relevant outputs, to end-users, so that end-users can critically examine and incorporate those outputs into their decision-making process, to the extent that they are deemed a correct interpretation of the specific context applicable to the end-users.

4. *Verification*. Depending on the nature of the data, in relation to how often it may change, controllers should verify the correctness of personal data with the Data Subject before and at different stages of the processing. Accuracy checks are not a one-off exercise. Depending on the nature of the data collected and the likelihood that that data may vary over time, a continuous effort to ensure accuracy may need to be carried out. This may involve further data collection (overwriting older data with newer data, to prevent excessive data aging) and ensuring that Data Subjects are granted means by which to update their own data, through rectification requests, among other measures.

5. *Erasure/rectification*. Controllers must erase or rectify inaccurate data without delay. Inaccurate data presumably bring no value to the policymaking process and may in fact be harmful towards that process. The analytic-ingest functions of the platform, as seen in Section 3.1.3.1 below, contribute towards this element. End-users will be responsible for configuring constraints and rules for the data cleaning process properly, as well as monitoring data sources used over time to ensure their continued accuracy, reliability, and relevance. Whenever a data source is deemed to no longer fit those requirements, it should no longer be used; this applies also to specific data points collected, or outputs built upon inaccurate data points.

6. *Accumulated errors*. Controllers must mitigate the effect of an accumulated error in the processing chain. This speaks to the importance of detecting inaccuracies or bias in underlying data used for output generation on the platform. Errors in the dataset may lead to errors in the policies ultimately developed.

7. *Access*. Data Subjects should be given an overview and easy access to personal data to control accuracy and rectify as needed. This implies that, whenever personal data is stored in an identifiable form of the platform, it should be possible to query those data to identify, extract and modify and/or delete personal data pertaining to an individual Data Subject, so that Data Subjects can effectively exercise their rights to rectification regarding inaccurate or incomplete personal data, under Art. 16 GDPR.

8. *Continued accuracy*. Personal data should be accurate at all stages of the processing, and tests of accuracy should be carried out at critical steps. This requires a combination of responsibility of the end-users for ensuring the selection of reliable data sources with the mechanisms of the platform for data cleaning and validation.

9. *Updated data*. Personal data shall be updated if necessary, for the purpose. In particular, the older data is, the greater the likelihood that it becomes obsolete. To avoid data aging, end-users should periodically revise whether data relied on remains accurate and relevant, refraining from using any data which no longer accurately reflects the underlying reality. Whenever a Streaming data source is used, appropriate rollover periods (i.e., periods after which older data is to be overwritten by newer data collected) should be defined.

10. *Data design*. Technological and organisational design features shall be used to decrease inaccuracy. Data collected from data sources should be properly categorised and classified, according to the specifications defined by end-users, to reduce the likelihood of errors in interpretation and increase effectiveness of the analytical processes to be applied on those data.

End-users, in collaboration with PolicyCLOUD, should carefully assess the technical and organisational measures which can feasibly be implemented to address all the above requirements, to ensure compliance with the principle of accuracy.

2.2.4.8 STORAGE LIMITATION PRINCIPLE

PolicyCLOUD shall keep personal data in a form which permits identification of Data Subjects for no longer than is necessary for the purposes for which the personal data are processed.

This means that even where personal data are collected in a fair and lawful manner, they cannot be stored for longer than actually needed, unless a reason for further processing exists, and provided that a legal basis for such further processing has been detected by PolicyCLOUD pursuant to the purpose limitation principle.

In light of this principle, PolicyCLOUD shall proceed to the erasure of personal data from the cloud-based platform when it has no reasons for keeping them or, alternatively it shall anonymize and aggregate such data, since anonymized data are no longer considered personal data within the meaning of the GDPR. It is therefore important for PolicyCLOUD to identify and implement appropriate data anonymization techniques which guarantee that relevant datasets are no longer referable to an identified or identifiable natural person.

This principle becomes practically relevant whenever personal data is collected, if it is not promptly anonymised or aggregated, with the underlying raw data being deleted as soon as possible. Noting that personal data avoidance is the preferred approach under the data minimisation principle, whenever this is not feasible, then end-users, as controllers, should define specific retention periods for the personal data collected, based on the strict minimum period of time for which retention of those data is needed to ensure that the purpose for their collection and

processing can be met. When dealing with streaming data sources, it is relevant to define an appropriate retention and/or overwrite period to avoid data aging (i.e., defining a short rollover period after which older data will be overwritten by newer data collected through the stream). The platform should allow the end-user to define retention periods and include tools for automatic deletion or aggregation of underlying data after an end-user-defined retention period is exceeded.

The key design and default elements to be considered, for compliance with this principle, as well as by the platform in its configuration, include [38]:

1. *Deletion and anonymisation*. Controller should have clear internal procedures and functionalities for deletion and/or anonymisation. This requires end-users to define clear internal retention policies for the personal data they intend to collect; the platform should be configurable to allow the retention requirements defined by the end-users to be effectively implemented on any stored personal data.

2. *Effectiveness of deletion and anonymisation*. Controllers shall make sure that it is not possible to re-identify anonymised data or recover deleted data and should test whether this is possible. This speaks to the effectiveness of the procedures put in place on the platform to aggregate, anonymise and delete personal data. The platform must be configured to ensure that these processes are effective and irreversible. [18]

3. *Automation*. Deletion of personal data should be automated. This element is also mostly upon the platform, which should allow for the implementation of the retention policies of the end-users without need for active intervention of the end-user.

4. *Storage criteria*. Controllers shall determine what data and length of storage is necessary for the purpose. End-users are responsible for defining the retention periods they deem appropriate, based on the strict minimum amount of time for which data must be retained in an identifiable form to allow the purpose to be met.

5. *Justification*. End-users shall be able to justify why the period of storage is necessary for the purpose and the personal data in question, and be able to disclose the rationale behind, and legal grounds for the retention period.

6. *Enforcement of retention policies*. Controllers should enforce internal retention policies and conduct tests of whether those policies are effectively followed. Both end-users in terms of defining the policies and PolicyCLOUD in terms of providing the means for the policies to be implemented will need to collaborate to ensure this element is achieved.

7. *Backups/logs*. Controllers shall determine what personal data and length of storage is necessary for backups and logs. The definition of retention periods should also consider the need to retain backups of personal data, for security and data availability purposes. These should be configured by PolicyCLOUD, with input from end-users, where feasible and/or relevant.

8. *Data flow*. Controllers should beware of the flow of personal data, and the storage of any copies thereof, and seek to limit their temporary storage. Temporary storage also refers to caching or redundant copies of personal data, for operational or security purposes: such copies should be limited as much as technically possible to ensure proper functioning of the platform, avoiding the creation of unnecessary copies to mitigate the risk of personal data breaches.

A joint effort from end-users and PolicyCLOUD is needed to carefully assess the technical and organisational measures which can feasibly be implemented to address all the above requirements, to ensure compliance with the principle of storage limitation. This is without prejudice to the preferred approach which, as mentioned in

Section 2.2.4.6 above, is to limit the storage of personal data as much as possible, avoiding storage altogether, or promptly aggregating/anonymising stored data, whenever feasible.

2.2.4.9 Integrity and Confidentiality (Security)

The principle of security laid out in Art. 5, par. 1, let. f) GDPR, in combination with Art. 32 GDPR, is a key reflection of the risk-based approach of the GDPR. Controllers and processors alike are required to ensure that any personal data processed are supported by appropriate security measures, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage. This should be done by defining and implementing appropriate technical and organisational security measures, considering the available technology (including the state of the art and the costs of implementation), the circumstances under which personal data are processed and the risks which may result to the rights and freedoms of individuals, and particularly those which may result from the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data, with the end-goal of ensuring a level of security appropriate to those risks.

To define appropriate security measures, there must first be an assessment of "*the risk of varying likelihood and severity for the rights and freedoms of natural persons*"[39] inherent to the processing activities to be performed. As the PolicyCLOUD platform is intended to be used in a variety of use cases, involving a variety of different data sources and data, it is not a simple matter to abstractly determine what level of security measures might be appropriate to comply with this principle. The recommended starting point would be to rely on the specific use cases currently being considered to identify an appropriate security baseline for future use cases. It should be possible to demonstrate that security measures chosen have been deliberately and cautiously selected, to address specific and identified risks for data subjects, in compliance with the principle of accountability – see Section 2.2.4.10, below.

The risk-based approach of the GDPR offers freedom and flexibility in deciding on appropriate security measures, but also creates uncertainty as to whether or not the implementation of particular measures may lead to a "*level of security appropriate to the risk*", as established in Art. 32, par. 1 GDPR. In practice, even the performance of a comprehensive risk assessment will typically not provide complete certainty as to the recommended or best means to address any security risks identified. To provide further assurance, internationally recognised information security standards, such as those of the ISO/IEC 27000 family, can be considered: however, alignment with those standards alone is not a guarantee of compliance with Art. 32 GDPR, as the specific processing activities carried out may generate particular risks which those standards do not address.

In this respect, existing guidance on security measures can support this decision-making process. As an example, the ENISA has developed guidelines aimed at digital service providers [41], which identify twenty-seven different security objectives, listing technical and organisational security measures for each objective, which are ranked in three different sophistication levels:

1. Level 1 reflects basic security measures, which may be implemented to reach the objective in question.
2. Level 2 reflects industry standard security measures, which not only allow the objective to be reached, but also the review of the implementation of that objective, in the event of relevant changes or incidents.

[39] Art. 32, par. 1 GDPR.

3. Level 3 reflects the state of the art, which are advanced security measures allowing for continuous implementation monitoring and structural implementation review, considering relevant changes, incidents, tests, and exercises, to proactively improve the implementation of those measures.40

Where feasible, it is strongly recommended for the platform to be configured in accordance with sophistication level 3, as ensuring a higher baseline for security measures increases the likelihood of the chosen measures being deemed compliant for the processing of a wider variety of personal data.

Another example which can be considered are the guidelines developed by the CNIL. [42] These guidelines point out basic precautions which should be systematically implemented to managing security risks involved in the processing of personal data, including:

1. The raising of user awareness on privacy and security challenges of each organisation.
2. The management of data and system access rights assigned to users, including the definition of those rights in a manner which ensures effective compliance with the principle of data minimisation, and the logging of access to personal data. This is currently addressed by D3.1. [43]
3. The management of security incidents and personal data breaches.
4. Measures which can be implemented to secure workstations, mobile equipment, internal networks, servers, and websites.
5. Backup policies and secure data archiving.
6. The performance of maintenance on data processing systems and the secure destruction of data.
7. The management of processors and transmissions of data to other organisations.
8. The physical security of premises.
9. Data protection by design and by default.
10. Measures to ensure the integrity, confidentiality, and authenticity of personal data.

Controllers and processors will be held accountable for their decisions related to security measures in the event of an inspection by a supervisory authority. Therefore, it should be demonstrable that the security measures implemented on the platform were chosen as a result of a documented risk assessment, with justifications as to why those measures were deemed adequate to address the specific risks identified.

The management of personal data breaches is a key component to be addressed by security measures implemented on the platform. "Personal data breach" is defined, under Art. 4, par. 12 GDPR, as "*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*": personal data breaches are security incidents which have a relevant impact on personal data. [44] Art. 32 GDPR specifically requires risks arising from the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data to be considered in

[40] Since these guidelines were drafted in 2017, it should be noted that the 'state of the art' is likely to have evolved since.

the definition of security measures. Recital 85 GDPR further highlights that "[a] *personal data breach may, if not addressed in an appropriate and timely manner, result in physical, material or non-material damage to natural persons such as loss of control over their personal data or limitation of their rights, discrimination, identity theft or fraud, financial loss, unauthorised reversal of pseudonymisation, damage to reputation, loss of confidentiality of personal data protected by professional secrecy or any other significant economic or social disadvantage to the natural person concerned*". As such, the platform should also be equipped with technical and organisational measures to ensure the ability to prevent and detect personal data breaches, as well as react to occurred breaches in a timely and compliant manner.

The concept of personal data breach is quite vast. Broadly speaking, personal data breaches can be classified as:

1. Confidentiality breaches, where there is an unauthorized or accidental disclosure of, or access to, personal data.
2. Integrity breaches, where there is an unauthorized or accidental alteration of personal data.
3. Availability breaches, where there is an accidental or unauthorized loss of access to, or destruction of, personal data. [44]

In practice, personal data breaches may range from acts of mere human error to acts of malicious interference with processing systems. The implementation of internal policies and procedures to ensure effective management of personal data breaches is therefore strongly recommended, alongside other security measures set out to manage breaches, aiming at the following overall objectives:

1. The detection of relevant security incidents.
2. The assessment of relevant security incidents, in terms of whether they may qualify as a personal data breach, and in terms of the severity of their impact to the rights and freedoms of Data Subjects affected.
3. The notification to the relevant supervisory authority and communication to Data Subjects, where relevant.
4. The documentation of personal data breaches managed.
5. Review.

There should be clearly defined rules and specific channels on the reporting of security incidents or abnormal events related to the platform. All persons working with the platform should be made aware of the types of occurrences which may qualify as a reportable security incident. Irregularities can also be detected through technical measures. [44] A team of competent individuals, including preferably the DPO [45] of the platform and members of the information and physical security departments, should be identified to address reports and manage any event qualifiable as personal data breach. In particular, this team will need to assess the extent to which it is required, under the GDPR, to notify the personal data breach to a relevant end-user acting as controller and/or to a competent supervisory authority or the Data Subjects affected, where the circumstances of Arts. 33 and/or 34 GDPR are met, for personal data for which PolicyCLOUD acts as a controller. That team will further be responsible for establishing appropriate mitigation measures to reduce identified risks and damages.

The obligations of PolicyCLOUD concerning a personal data breach, in terms of notifications/communications under the GDPR, will vary depending on the data protection role taken by PolicyCLOUD regarding the affected personal data:

1. Where acting as a controller (i.e., where the data affected is processed by PolicyCLOUD for its own, independent purposes, such as personal data collected on individual users of the platform), PolicyCLOUD is required, under Art. 33, par. 1 GDPR, to report any personal data breach detected to a competent supervisory authority [44] within seventy-two hours of becoming aware of the breach [44] unless the personal data breach in question is deemed unlikely to result in a risk to the rights and freedoms of individuals. [44] Art. 33, par. 2 GDPR describes the minimum content which these notifications should include[41] Where it is not possible to provide all required information within the first seventy-two hours, all relevant information at disposal of PolicyCLOUD should be provided within that deadline, updating the notification with additional details as they become available ("notification in phases"). [44] While it is generally preferable to follow the notification in phases approach in these cases, it is also possible, exceptionally, to delay the first notification beyond this deadline [44], as long as this can be reasonably justified to the supervisory authority. If the assessment performed by PolicyCLOUD of the severity of a personal data breach indicates a high level of risk to the rights and freedoms of individuals, then Art. 34 GDPR will require PolicyCLOUD, as a rule, to directly inform the affected individuals of the occurred breach, without undue delay.

2. When acting as a processor, PolicyCLOUD is only required to communicate detected personal data breaches to the relevant controller (i.e., the end-user), under Art. 33, par. 2 GDPR. In this context, there is no requirement to carry out a risk assessment pertaining to the personal data breach; rather, once it has been established that a personal data breach has occurred, the appropriate controller must be informed without undue delay. [44] PolicyCLOUD will also be required to further cooperate with the relevant controllers to further investigate and collect information on the personal data breach in question.

The processes implemented to address personal data breaches by PolicyCLOUD should also ensure that all relevant information on a personal data breach and the manner in which it was handled is documented in a register of personal data breaches, as set out in Art. 33, par. 5 GDPR, including all facts pertaining to the personal data breach, its effects and remedial action taken (including notifications to end-users, supervisory authorities and/or Data Subjects, as well as all technical and organizational mitigation measures applied), documented assessments carried out (including those performed to classify the incident as a personal data breach, as well as to classify a personal data breach in terms of category and severity level). Post-breach analyses should also be carried out, to validate the effectiveness of the breach management process, identify areas of improvement, and identify, based on a root cause analysis of the incident, adequate technical and organisational measures to reduce or eliminate the likelihood of recurrence.

It is fundamental for end-users to ensure that the security measures offered by the platform allow for an adequate level of data security, considering the potential risks for data subjects inherent to the processing of personal data in the context of their use cases. As such, a security risk assessment, as part of an overall personal data protection impact assessment, should be carried out, to identify possible threats and risks to the fundamental rights, freedoms

[41] The notification shall include a description of the nature of the personal data breach (including, where possible, the categories and approximate number of data subjects concerned, and the categories and approximate number of personal data records concerned), the name and contact details of the DPO of the controller or other point of contact, a description of the likely consequences of the personal data breach (as assessed by the controller), and a description of the mitigation measures taken by the controller to address the breach or those which the controller proposes to be taken.

and interests of Data Subjects and the specific security measures implemented or which should be implemented to address them.

2.2.4.10 ACCOUNTABILITY PRINCIPLE

Accountability shall be considered as the overarching principle which encompasses all the above-described principles PolicyCLOUD and end-users are subject to when processing personal data.

The accountability principle requires PolicyCLOUD and end-users to take responsibility for what is done with personal data and how it complies with the other principles set out above, implementing measures, documents and records to demonstrate that appropriate processes and procedures are in place to ensure that personal data are collected, processed and stored in such a way that is compliant with the GDPR and with other applicable data protection laws.

Accountability is therefore an ongoing exercise and relates to the need for PolicyCLOUD and end-users to keep track of processing activities carried out for the execution of the Project and be able to motivate the choices made with regard to personal data processing. From this principle derives the obligation for PoIicyCLOUD and end-users to prepare a series of documents, policies and procedures required under the GDPR, including:

1. A record of processing activities carried out by PolicyCLOUD and end-users in their capacity as controllers, which shall contain at least:
    - Information on the name and contact details of the controller and, where applicable, the joint controller, the representative of the controller and/or the DPO.
    - The purposes of the processing.
    - A description of the categories of Data Subjects and of the categories of personal data processed; the categories of recipients to whom the personal data have been or will be disclosed, including recipients in third countries or international organisations.
    - Where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation.
    - Where possible, the envisaged time limits for erasure of the different categories of data.
    - Where possible, a general description of the technical and organisational security measures implemented to ensure a level of security of personal data appropriate to the envisaged risks personal data may be subject to.
2. A procedure for the management of data breaches and a related record for keeping track of such events.
3. A procedure for managing possible requests coming from Data Subjects who want to exercise their data protection rights granted by the GDPR, along with a related record for keeping track as to when and how such requests were addressed.
4. To the extent that consent is relied on as a legal basis for the processing of personal data, adequate records of consent provided by Data Subjects, meeting all applicable requirements of validity under the GDPR.
5. To the extent that legitimate interests is relied on as a legal basis for the processing of personal data, a documented legitimate interests assessment, to demonstrate that data subjects' fundamental rights, freedoms and interests are not disproportionately impacted by the intended processing activity.

6. The drafting of specific privacy policies.

7. The drafting of a DPIA each time, considering the type of processing, in particular when using new technologies, as well as the nature, scope, context and purposes of the processing, PolicyCLOUD deems likely that a processing activity would result in a high risk to the rights and freedoms of data subjects.

8. For PolicyCLOUD, detailed records on its relationship with providers which process personal data on behalf of PolicyCLOUD (thus acting in their capacity as data processors), to present evidence of due diligence and compliance in terms of evaluation of engaged providers as to technical and organizational security measures they have in place to protect personal data processed in the context of the Project and of overall compliance with the data protection legal framework, as well as DPAs entered into with such providers.

9. For end-users, given that PolicyCLOUD is arguably acting as a processor on behalf of the end-user regarding any processing of personal data carried out at the specifications of the end-user, to allow the end-user to benefit from the services provided by PolicyCLOUD, it will be important for end-users to enter into and document an appropriate DPA, under Art. 28 GDPR, with PolicyCLOUD, and to archive this agreement.

2.2.5 Future regulatory issues: the DGA

From a mid and long-term perspective, since PolicyCLOUD will use analytics technologies using large datasets to enable policymakers to undertake their decisions, it will be of paramount importance to monitor the impacts related to the possible approval of the DGA.

Indeed, the EC published its DGA proposal on November 25, 2020. It is one of several incoming pieces of legislation proposed at the EU level, to accomplish the European Strategy for Data, adopted in February 2020, and create an EU single market for data.

The DGA will introduce:

1. Rules for making public sector data available for reuse, in situations where such data is subject to rights of others.

2. A framework allowing companies to share industrial data in common data lakes, called European Data Spaces.

3. Rules for data brokers, called “data sharing providers,” including a notification regime and the obligation to remain neutral as to the data exchanged.

4. The concept of “data altruism,” allowing people to share data for the common good.

The bill still needs to be approved by both the European Parliament and the Council of the European Union. [46]

3 Ethical, legal, societal, and regulatory issues related to PolicyCLOUD components

In this section, we will further develop the abovementioned requirements, from the ethical, legal/regulatory and societal perspective, by more specifically setting out the particular issues raised by the components of the Project and the activities performed on the data (in an abstract manner, without reference to any specific use case). Other than the relevant personal data protection and privacy laws mentioned above, a strong emphasis will be placed on recommendations provided by relevant European and international institutions, with the end-goal of further specifying the requirements identified in section 1 above for the use of analytical tools.

3.1 Cloud Capabilities & Data Collection Engine

3.1.1 Use of Cloud Infrastructure

3.1.1.1 Contractual regulation of data processing relationships

As noted in D3.1 [43], the PolicyCLOUD project will be reliant on a cloud-based infrastructure to provide for the necessary computing and storage capabilities to allow the project to fulfil its goals. This infrastructure will be provided by an IaaS provider, with such provision currently regulated under the contractual combination of an Operational Level Agreement and an SLA.

Where the provision of this infrastructure may involve the processing of personal data on behalf of PolicyCLOUD, the IaaS provider may be acting as a processor on behalf of PolicyCLOUD. In this scenario, Art. 28, par. 3 GDPR requires the processing activities performed by the IaaS provider to be governed by "*a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller*". This is typically referred to as a DPA. Art. 28, par. 3 GDPR goes on to establish a set of minimum obligations which must be included in such a data processing agreement, which have been further refined by the EDPB [47]:

1. The processor must process the personal data only on documented instructions from the controller, including with regard to transfers of personal data to a third country or an international organisation, unless required to do so by EU or member State law to which the processor is subject. In such a case, the processor shall inform the controller of that legal requirement before processing, unless that law prohibits such information on important grounds of public interest.[42]

[42] See Art. 28, par. 3, let. a) GDPR. Where a processor seeks to further process personal data for its own purposes, outside of the instructions of the controller, that processor may be qualified as a controller for such further processing, under Art. 28, par. 10 GDPR. These instructions will typically be provided within the DPA and the connected service agreements, though it is possible for the controller to issue ad hoc instructions to a processor on the use of personal data, unless stipulated otherwise between the parties.

2. The processor must ensure that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality.[43]

3. The processor must implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons.[44]

4. The processor must not engage another processor without prior specific or general written authorisation of the controller.[45] Furthermore:

    - Where a general written authorisation is granted, the processor must inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes. The processor must impose the same data protection obligations as set out in the data processing agreement with the controller on that other processor, by way of another agreement, which can be called a "data sub-processing agreement".

    - The processor must remain fully liable to the controller for the performance of the obligations of other processors engaged.

5. The processor must assist the controller, by appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the obligation of the controller to respond to requests for exercising the rights of the Data Subjects, taking into account the nature of the processing.[46]

6. The processor must assist the controller in ensuring compliance with the obligations of the controller around security of personal data, notification of personal data breaches to supervisory authorities, communication of personal data breaches to Data Subjects, DPIAs and prior consultations.[47]

[43] See. Art. 28, par. 3, let. b) GDPR. "Persons authorised to process personal data" is a broad concept and includes processor employees as well as temporary workers.
[44] See Arts. 28, par. 3, let. c) and 32 GDPR. Data processing agreement should specify the security measures which the processor commits to have implemented, with the processor being obliged to obtain the approval of the controller before making changes to those measures.
[45] See Arts. 28, par. 2, 28, par. 3, let. d), 28, and 28, par. 4 GDPR.
[46] See Art. 28, par. 3, let. e) GDPR. The responsibility for addressing the requests of the Data Subjects rests primarily on the controller, under Art. 12 GDPR. However, processors are required to provide reasonable assistance in this respect. This assistance may range from simply relaying any requests received to the controller, to more specific, technical duties around response, where the processor is able to extract and manage personal data. The processor should be given specific instructions on how to proceed when faced with a data subject request.
[47] See Arts. 28, par. 3, let. f) and 32 to 36 GDPR. In particular, the processor must: 1. assist the controller in ensuring an appropriate level of data security; 2. promptly notify personal data breaches detected to the controller and help gather further information on such breaches, and 3. provide information on its processing activities as reasonably needed to allow the controller to perform DPIAs or requests for prior consultation from a supervisory authority.

7. The processor must, at the controller's choice, delete or return all the personal data to the controller after the end of the provision of services relating to processing, and delete existing copies unless Union or Member State law requires storage of the personal data48.

8. The processor must make available to the controller all information necessary to demonstrate compliance with the above obligations and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller49.

Given that several of the data sources identified as relevant for the PolicyCLOUD project in, e.g., deliverables D2.1 [48] and D6.3 [49], may contain personal data (e.g., social network posts and content, websites and blogs), and that certain scenarios identified for the different use cases (as covered in Sections 4 to 7 below) may require the capturing and further storage or hosting of personal data (potentially in an identifiable, non-aggregated form) on this cloud infrastructure, it is reasonable to maintain that the provision of a cloud infrastructure will necessarily involve activities which can be qualified as personal data processing activities, performed by the IaaS provider as a processor on behalf of PolicyCLOUD.[50]

This conclusion is further supported by the WP29 Opinion 05/2012 on Cloud Computing [50], as the WP29 notes that, as a rule, "[w]*hen the cloud provider supplies the means and the platform, acting on behalf of the cloud client, the cloud provider is considered as a data processor*" (p. 8), and the IaaS model is specifically identified as within scope of this Opinion (p. 26). It is also supported by the EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR [47], given that, according to the existing SLA and Operational Level Agreement, the IaaS provider will be processing (storing and/or hosting) data for the benefit of PolicyCLOUD, outside of the direct authority of the control of PolicyCLOUD, and only for purposes defined by PolicyCLOUD (i.e., the IaaS provider will not reuse PolicyCLOUD data for its own purposes). The storage and hosting functions provided by the IaaS provider are a key element of its services, and the IaaS provider retains autonomy in deciding most technical and organisational aspects as to how those functions will be provided, while not being able to make essential decisions as to how the stored data will be processed (e.g., defining the types of data stored, the retention periods applied to the data, the rights to access the stored data), given that the IaaS provider will presumably not have access to the stored data in itself. Indeed, these decisions will remain with PolicyCLOUD, as the controller.

As such, to the extent that personal data may be processed through the PolicyCLOUD platform, which is built upon the cloud infrastructure provided by the IaaS provider, compliance with Art. 28 GDPR will arguably require the

[48] See Art. 28, par. 3, let. g) GDPR. It is up to the controller to decide, within the DPA, whether personal data held by the processor should be deleted or returned upon termination of services. Any deletion should be performed in a secure manner and confirmed to the controller within an agreed timeframe.
[49] See Art. 28, par. 3, let. h) GDPR. The DPA should specify how information will flow between the controller and the processor, so that the controller can remain fully informed as to the details of the processing activities of the processor. For example, the processor may disclose its record of processing activities to the controller developed under Art. 30, par. 2) GDPR), provide periodic reports to the controller, or otherwise make itself available to address any specific inquiries which the controller may have over time. The GDPR also explicitly requires processors to make themselves available for audits performed by or on behalf of the controller, and so the DPA should regulate the terms under which such audits may be carried out.
[50] Note that the definition of "processing" provided by Art. 4, par. 2) GDPR is very broad and includes the mere storage of personal data.

execution of a DPA between PolicyCLOUD and the IaaS provider[51], including at least all of the obligations mentioned above.

3.1.1.2 Cloud Security and Compliance

It is important for PolicyCLOUD, not only from the legal, but also from the ethical perspective, to ensure the security and resilience of its platform, so as to preserve the availability of data, policies, algorithms and other assets contained on the platform, as well as of the platform itself. Furthermore, the risks around obstacles created by use of a cloud-based infrastructure to compliance with applicable legal obligations (e.g., related to personal data protection) should also be addressed.

It is therefore vital for PolicyCLOUD to:

1. Perform a comprehensive risk assessment, focused on the likelihood and impact of threats to confidentiality, integrity and/or availability of assets stored on the platform, and to the resilience of the infrastructure of the platform and systems itself, as well as relevant compliance obstacles raised by use of the cloud infrastructure. [51]
2. Assess whether the technical and organisational measures put in place by the IaaS provider in relation to the cloud infrastructure sufficiently mitigate any relevant risks identified.

In performing this assessment, it is recommended that PolicyCLOUD consider, at least, the following risks:

1. Loss of governance regarding the use of data stored on the cloud infrastructure.
2. Technological dependency on the IaaS provider, such that changing solutions or providers becomes overly burdensome or implies loss of data.
3. Flaws in data isolation, allowing unauthorised access to data stored on the cloud infrastructure.
4. Data disclosure requests which may be filed by competent public authorities against the IaaS provider.
5. Flaws in the subcontracting chain, to the extent that the IaaS provider uses third parties to provide or support the cloud infrastructure.
6. Ineffective or insecure data deletion, or excessive data retention periods defined by the IaaS provider.
7. Impossibility to properly govern data access rights on the cloud infrastructure.
8. Unavailability or reduced availability of service, including the infrastructure itself and access to the infrastructure.

[51] Given the model of cloud infrastructure provision, there will be no direct contractual relationship between PolicyCLOUD and the IaaS provider. To ensure that a chain of data processing agreements is properly established, in compliance with Arts. 28, par. 3 and 4 GDPR, it is recommended that PolicyCLOUD (as a controller) enter into a data processing agreement with EGI (as a processor), and that EGI (as a processor) enter into a data sub-processing agreement with the IaaS provider (as a sub-processor).

9. Shutdown of service, or takeover of service by a third party.
10. Failures on the IaaS provider's part to comply with applicable laws and regulations related to the service.
11. Cross-border transfers of data carried out by the IaaS provider without appropriate regulation under applicable laws (particularly where personal data may be transferred).
12. Where personal data may be stored, lack of proper configuration of the service to allow, e.g., requests for erasure, restriction of processing or portability to be promptly and properly complied with.
13. Changes to service terms without prior consent of PolicyCLOUD.
14. Further use of data stored on the service for purposes not authorised by PolicyCLOUD.

In assessing the IaaS provider from the security perspective, it is recommended that PolicyCLOUD guides itself according to existing standards, as these may provide credible and authoritative support in determining whether the IaaS provider's security measures may be deemed adequate. It will thus be important for PolicyCLOUD to determine what standard (or standards) to rely on. To this end, examples which PolicyCLOUD may wish to consider include the Information Assurance Framework developed by the ENISA [52], the Cloud Controls Matrix developed by the Cloud Security Alliance [53], or relevant ISO/IEC standards.[52]

3.1.2 Data Collection

3.1.2.1 Compliance Assessment around Selection of Data Sources

As noted in Section 2.2 above, applicable legal, regulatory, ethical and/or societal requirements may restrict an end-user's leveraging of a data source which might otherwise be considered appropriate from a practical perspective (i.e., useful towards the goal which the PolicyCLOUD end-user wishes to achieve).

It is important to assess whether each of the data sources currently envisaged to be used by PolicyCLOUD (for each of the use cases, as noted in Sections 4 to 7 below), can be used in compliance with such requirements (e.g., whether the data contained within the data source may be collected lawfully, considering the further processing to which it will be subjected within the platform and the end-goal for which it is collected by the PolicyCLOUD end-user). As a result of these assessments, specific data sources may need to be excluded, or be restricted in terms of the amount or type of information extracted from them.

In particular, the following points must be borne in mind during these assessments:

1. Should a selected data source be subject to contractual terms which prevent its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the data source owner (i.e., the person or organisation entitled, under the contractual terms, to authorise use of the data source).
2. Should a selected data source be eligible for database copyright or sui generis right protection which prevents its use as intended by the PolicyCLOUD user, that data source should not be registered on

[52] Including ISO/IEC 27001:2013, ISO/IEC 27002, and ISO/IEC 27017:2015.

PolicyCLOUD without proper authorisation from the rights holder (i.e., the person or organisation entitled, under the law applicable to the data source, to authorise use of the data source).

3. Should a selected data source, or a relevant part of that data source, which is to be extracted, be eligible for copyright which prevents its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the rights holder.

4. Should a selected data source contain personal data (i.e., information which can reasonably be linked to specific individuals), that data source should not be registered on PolicyCLOUD without:

   - Identifying one or more specific purposes for which that data source is to be used.
   - Identifying a valid legal basis, under Art. 6 GDPR, and a valid derogation, under Art. 9 and/or Art. 10 GDPR, if applicable, for each of those purposes.

5. Should a data source, containing personal data, be selected for PolicyCLOUD development purposes (i.e., as training data), that data source should not be registered on PolicyCLOUD without the completion of a legitimate interests assessment which demonstrates that the interests of PolicyCLOUD in using that data source for those purposes are not overridden by the interests and fundamental rights or freedoms of the corresponding Data Subjects, as well as, where relevant, the identification of appropriate derogations under Art. 9 and/or Art. 10 GDPR.

6. Should a data source, containing personal data, be selected for PolicyCLOUD deployment purposes (i.e., use of the PolicyCLOUD platform by an end-user), that data source should not be registered on PolicyCLOUD without the PolicyCLOUD user having identified a specific purpose for use of that data source, and demonstrating that a corresponding legal basis and, where applicable, derogations under the GDPR which justify that use have been identified, or at least warranting that such has been done, under their responsibility.

7. Should a selected data source contain personal data (i.e., information which can reasonably be linked to specific individuals), that data source should not be registered on PolicyCLOUD without:

   - Mapping out the specific data points which are to be extracted from that data source.
   - Identifying a specific justification for each data point, based on a strict need to use that data point to meet the purpose for which the data source was selected.

   Data points which are not strictly needed should not be collected. Whenever possible, direct, or indirect identifiers should not be collected, and any personal data collected should be aggregated or anonymised as soon as feasible.

8. Should a selected data source contain personal data (i.e., information which can reasonably be linked to specific individuals), that data source should not be registered on PolicyCLOUD without an assessment as to the likelihood and impact of the output generated by PolicyCLOUD based on that data source leading to bias or discrimination against individuals and, whenever feasible, identifying specific mitigation measures which can be implemented to address those risks.

9. Any selected data sources must be appropriately vetted in terms of their reliability, so as to prevent the collection of false, inadequate, inaccurate or incomplete data, whether personal or not, which might lead to erroneous or misleading output generation via PolicyCLOUD. The risk that false or manipulated data might be introduced into the platform to abuse or manipulate policy-making processes must be considered during this vetting process.

For future data sources (i.e., data sources which are later incorporated into the platform, by future PolicyCLOUD end-users), it is reasonable to maintain that this assessment should be performed by the end-user (given the plurality of possible data sources and the complexities around this type of assessment, which may create practical difficulties in automating the assessment on the side of PolicyCLOUD). The obligation to ensure the possibility for lawful use of any data source to be registered, including the need to obtain any potential licenses or authorisations from rightsholders over such data sources, would need to be expressly set out for PolicyCLOUD end-users in the terms and conditions governing use of the platform (see Section 3.3., below), to limit the liability of PolicyCLOUD on this matter to the greatest extent permissible by the applicable law.

3.1.2.2 INCENTIVES MANAGEMENT

As noted in D3.1 [43], there is currently not a great level of detail defined for the incentives management component of the PolicyCLOUD platform.

However, to the extent that this component seeks to collect input from individuals (i.e., citizens) to assist in the policy-making process via the platform, concerns around the protection of personal data related to participating individuals may be triggered. In particular:

1. *Lawfulness*. An appropriate legal basis for use of input provided by individuals must be defined, under Art. 6 GDPR, if that input amounts to, or contains, any personal data related to those individuals. One particularly relevant legal basis for such activities may be the use of consent. Where this is the determined approach, all requirements for valid consent established under the GDPR must be met[37]

2. *Fairness*. Input provided by individuals should not be used in a manner which may violate their reasonable expectations. Individuals should not be discriminated against or targeted with an aim to cause them harm due to having provided feedback in connection with a specific policy-making process. There should also be mechanisms in place to ensure that individuals are able to exercise their rights in relation to any personal data of theirs which may be collected during the input provision process (e.g., right to erasure, right to withdrawal of consent, right to rectification).

3. *Transparency*. Individuals must be provided written information as to how their personal data may be handled as a result of their participation, in a concise, transparent, intelligible and easily accessible form, using clear and plain language, meeting all of the requirements of Arts. 13 and 14 GDPR. [39] This can be done, e.g., by developing an appropriate information notice / privacy policy to be shared with those individuals, prior to their participation.

4. *Purpose limitation*. Personal data contained within input submitted by individuals should not be used or shared with third parties for any purpose which is incompatible with the policy-making process.

5. *Data minimisation*. Only the strict minimum amount of personal data needed to properly process a feedback submission of an individual should be collected and further used. If it is possible to collect feedback anonymously (i.e., without any collection of personal data), this is the preferred option.

6. *Accuracy*. Individuals should be given the possibility to rectify any personal data they submit in connection with feedback provided (this is linked to the above-mentioned fairness principle).

7. *Storage limitation*. If personal data must be collected (in compliance with the above-mentioned data minimisation principle), then it may only be stored for the strict minimum amount of time needed to properly process a feedback submission of an individual, after which the personal data should be irreversibly anonymised or deleted. This will require the identification of appropriate retention periods for any such personal data collected.

8. *Security*. Appropriate security measures to ensure the confidentiality, integrity, and availability of any collected and further stored personal data, as well as the resilience of systems used to collect and further store those data, must be implemented.

9. *Accountability*. Records of evidence showing compliance with these requirements must be kept.

This section will be further developed in future versions of this deliverable, as further detail is made available on this component.

3.1.3 Data Preparation

3.1.3.1 Analytics-Ingest Functions

As explained by deliverable D4.1 [54], any data which is ingested into the platform from a selected and registered data source will go through an initial transformation and processing phase, before any further analytics activities are performed on those data. This will be performed by analytic-ingest functions" which, at present, are foreseen with several goals in mind, including:

1. Data cleaning, to address relevant privacy and security issues and/or to improve data accuracy.

2. Data filtering, to remove irrelevant data.

3. Data minimising, to remove data which may not be needed for further analytics.

4. Format transformation, to ensure that ingested data can be manipulated by all analytical components registered on the platform.

5. Alert generation, to notify the platform and/or the user when new data is collected from a streaming data source, as opposed to a data source which is fully ingested or merely queried.

As noted in Sections 2.2.4.7 and 3.1.2.1 above, whenever personal data are collected from a registered data source for further processing on the platform, it is necessary, under the principle of accuracy, to ensure the accuracy and completeness of any personal data collected to the greatest extent possible. From an ethical/societal perspective, as noted in Section 2.1.3 as well as Section 3.1.2.1 above, inaccurate or incomplete data may ultimately affect the output produced by the platform and used by PolicyCLOUD end-users to create policies, which may lead to unforeseen and unjustified harmful impact on individuals and communities. The data cleaning workflow, as described in D4.1 [54], should aim to meet the following goals:

1. Based on a set of data constraints and rules, the platform should seek to only operate on clean, correct, and useful data. As noted in Section 3.1.2.1 above, where a registered data source contains personal data, it should be technically possible to avoid the collection of direct or indirect identifiers, or of any other data points for which a specific justification has not been defined by the end-user, based on a strict need to use that data point to meet the purpose for which the data source was selected. Thus, the PolicyCLOUD end-user should be able to configure those constraints and rules to specify those data points which are strictly relevant and exclude all others. The default settings should require the end-user to identify specific data points to be collected (i.e., the default should be that no data are collected without an action from the end-user), rather than identify specific data points to be excluded (i.e., the default should not be that all data are collected, unless the end-user chooses otherwise).

2. After a dataset has been validated, the extracted data will be further assessed to address missing, irregular, unnecessary and/or inconsistent data. This is vital from the accuracy perspective: as mentioned above,

use of inappropriate data may impact the output of the platform, potentially leading to the creation of misguided policies. As this step is also based on data constraints and rules, these must be followed as defined by the end-user, so as to ensure that only justifiable data points (as defined at the previous point) have been collected, that those data points are useable and consistent, and that any collected data points which do not meet those requirements are removed.

3. End-users will be primarily responsible for selecting appropriate data sources, including in terms of their representativeness and potential for bias or discrimination in their composition. This must clearly be brought to the attention of the users of the platform attention, as the fact that data extracted from a data source has been validated and cleaned will not suffice to address this. This responsibility should be highlighted in the terms and conditions for the use of the platform, as noted in Section 3.3, below.

4. Logs should be kept, so that it is possible to trace back all actions taken by the platform on a specific data source, from collection to output, including all data cleaning operations performed. This is important to ensure the explainability of output provided by the platform in relation to the input received, and to demonstrate that end-user defined constraints and rules have been followed.

3.1.4 Enhanced Interoperability

As described by D4.1 [54], the Enhanced Interoperability component of the platform seeks to extract semantic knowledge and valuable information from the end-result of the data cleaning process. Whatever is extracted will become the input upon which further analytics may be performed, to produce an output useful towards policy development. This component will rely on standard vocabularies and classifications to categorize and sort data sources and data points, as well as linked data, natural language processing and semantic AI technologies and techniques to establish connections between data sources and data points.

In particular, the functioning of this component can be divided into two distinct steps:

1. The analysis, transformation, and annotation of cleaned data.

2. The interlinking of annotated data, through ontology mapping.

The accuracy and value of such correlations will depend not only on the accuracy of the underlying data, but also on the accuracy of the technologies and techniques used, i.e., on the accuracy of annotations and of the connections established between annotated data. Failing to establish appropriate connections may, at best, fail to provide fruitful additional information to the end-user and, at worst, actively provide misleading or erroneous information, culminating in misguided policy-making activities.

As such, to the extent that this component is to function on the basis of automated analysis, transformation, annotation and interlinking through AI systems, PolicyCLOUD should ensure that a sufficient amount of testing is performed to ensure a reasonable degree of statistical accuracy for annotations and connections established. This may require an extensive training exercise involving various kinds of data sources, to refine rules used by this component for these activities.

It is further recommended that this component methodically logs the activities performed to ensure traceability and presents correlations established and their rationale to the end-user for confirmation, to provide some level of human validation of those correlations. It is important that the end-user can understand the operations performed by the platform on the data; otherwise, the end-user may not be able to understand the connections established by the platform. Where this understanding is missing, the end-user may not feel confident in relying on the output of the platform or may be deprived of any ability to critically examine that output and ensure that it amounts to a correct interpretation of the context in which the data are being considered. End-users should also

be advised of the possibility of false positive correlations, so that they are incentivised to verify the validity of correlations made. Input from end-users can be used to further refine correlation-defining rules.

3.1.5 Data Storage

D4.1 [54] mentions the possibility, after the data cleaning process, for further storage of extracted knowledge on the platform, as opposed to further storage of the raw data points collected. Whenever personal data is collected, and to the extent that this can technically be achieved, this approach is preferable from the data minimization perspective. "Extracted knowledge" may potentially lose its qualification as personal data if the resulting information can no longer be linked to any specific individuals.

Where this is not possible, and there is an actual need to further store collected personal data within the platform, the principle of storage limitation set by the GDPR requires the definition of an appropriate retention period, based on the purpose for which those personal data were collected, and the strict need for their retention for a defined period for that purpose to be met. After this retention period, stored personal data would need to be deleted, or otherwise irreversibly anonymised or aggregated.

As such, whenever the end-user requires further storage of raw personal data, the end-user should be required to define a retention period after which such deletion, anonymisation or aggregation will take place. Given the possibility to transfer data to cold storage over time, this retention period could foresee a point prior to its end when such a transfer would take place.

The definition of a retention period for raw data collected from streaming data sources which is further stored on the platform must take into account the age of the data, allowing end-users to define the retention period as a roll-over period so as to overwrite older and potentially obsolete data with newer data collected over time.

3.2 Reusable Models & Analytical Tools Engine

3.2.1 Data Analytics

Following the abovementioned data cleaning process, carried out through the use of Analytic-Ingest functions, as well as additional steps to improve the semantic and syntactic interoperability of data and datasets through the Enhanced Interoperability component of the platform, different analytical tools may be applied to data ingested into the platform or, in the case of external data sources, to data stored externally which is queried. As of the date of this deliverable, the following specific analytic functions are foreseen to be built into the platform:

1. Situational Knowledge Acquisition & Analysis.
2. Opinion Mining.
3. Sentiment Analysis.
4. Social Dynamics.
5. Behavioural Data Analytics.

For all the above listed analytics functions, from an ethical and societal perspective, the concern as to sufficient statistical accuracy in the presentation of results is also present. A failure to provide appropriate analytical results may ultimately lead to misleading or erroneous conclusions drawn by policymakers, culminating in misguided policy-making activities. As such, PolicyCLOUD should ensure that enough testing is performed to ensure a

reasonable degree of statistical accuracy for these activities. Any operations performed on data should be methodically logged to ensure traceability, and the general rationale and logic behind the analytics performed should be explained to end-users, so this can be considered during their decision-making process. End-users should be advised of the possibility of false positives or false negatives and errors in result presentation, so that they are incentivised to critically examine results produced by the analytics functions in their decision-making process.

3.2.1.1 Situational Knowledge Acquisition & Analytics

This analytic function will rely on machine learning techniques to derive additional information from ingested data. As noted in D4.1 [54], this process will involve steps to categorise datasets (i.e., to structure analysed data into categories, so as to present text-based information to end-users on the data gathered) and to perform exploratory analysis on those categorised datasets (i.e., to allow end-users to visualise the main insights and features derivable from the structured categories).

As acknowledged in D4.1 [54], both of these steps must be defined for each particular use case, as the relevant categories, structure and method of information presentation will vary depending on data sources and the purposes for which they are used. End-users should thus be offered tools with which to define their requirements for structuring and presentation of the data extracted from registered data sources.

3.2.1.2 Opinion Mining

The information available on this analytic function, as described in particular in deliverable D2.2 [55],suggests that it is mostly aimed towards the collection of data from streaming data sources and in particular social networks, in order to identify and observe events and social attitudes in relation to specific topics and entities. The use of personal data for these activities is strongly indicated by the goal to be able to determine the main influencers or most popular users commenting on specific topics, which implies the collection of information on identifiable individuals.

From the personal data protection perspective, based on the limited information available, the following points should currently be borne in mind for such activities[53]:

1. *Lawfulness*. An appropriate legal basis, under Art. 6 GDPR, for the use of this function, where it involves analytics on personal data, must be identified and implemented by the end-user. In particular, the end-user may need to assess whether it can leverage its own legitimate interests as a legal basis, under Art. 6, par. 1, let. f) GDPR, by performing a legitimate interest assessment, or whether any other legal basis may apply.

2. *Fairness*. Only publicly available data should be considered for collection from social networks, as there is a lesser or possibly non-existent expectation of privacy around such data. Posts, comments, reactions, and other content uploaded onto private profile pages, groups or exchanged through direct messages should not be accessed or used.

3. *Transparency*. Measures to inform social network users about these activities must be taken, in compliance with Art. 14 GDPR. Where it is not feasible to contact users directly, given the sheer number

[53] These requirements will be subject to revision as this component is further developed.

of users involved, the PolicyCLOUD end-user must explore the possibility to rely on the exception under Art. 14, par. 5, let. b) GDPR. [39]

4. *Purpose limitation*. Information collected on social network users should not be used to cause them harm or in a manner which would violate their legitimate expectations.

5. *Data minimisation and storage limitation*. A strict temporal scope for data collection should be defined, to avoid data aging and excessive collection of information, which, if not recent, may not even be useful for the analytics activities aimed at by this function. A strict geographical scope for data collection should be defined, where feasible according to the specific use case.

6. *Rights of the data subjects*. Where data is stored in an identifiable form (i.e., linkable to specific social network users), it should be possible to query the data set and extract and delete all information collected and generated pertaining to individual users, within a short amount of time and in an effective manner, to ensure that their rights under the GDPR and, in particular, the right of access, the right to erasure and the right to object, can be satisfied.

3.2.1.3 Sentiment Analysis

The conclusions related to this analytic function are similar to those of the Opinion Mining function, as both of these may aim at collecting input from individuals on specific topics (in the case of Sentiment Analysis, policies developed by an end-user) from social networks. However, according to D2.2 [55], other channels for collection of this input, including channels provided or managed by end-users themselves, may also be considered, such that social networks may not be the primary focus of this function.

The similarities between the goals of this function and the activities related to the incentives management activities, in the sense that both seek to derive value from input provided by individuals related to policies developed or under development, suggest that the conclusions reached in Section 3.1.2.2 above should also be considered here.

3.2.1.4 Social Dynamics (Politika) and Behavioural Data Analytics

According to the descriptions provided by D2.2 [55] and D4.1 [54], these components refer to a web-based environment in which end-users will be able to create graph-based population models, according to user-defined parameters and relying on data extracted from data sources aligned with those parameters, to simulate the potential social effects of a given policy, by analysing the effects which may be felt on individuals and groups, i.e., connections between individuals.

To the extent that these components do not actually rely on information about identifiable individuals, but instead rely on appropriately aggregated and categorised data referring to generic and unidentified individuals and groups, which cannot be traced back to any given specific and identified individual, no specific personal data concerns arise. This is the preferred approach for these data visualisation tools: to provide end-user-facing results based solely on aggregated, non-personal data, or at least to provide those results in a form which does not reveal or disclose any personal data used to generate them (i.e., which does not provide specific information about identified individuals).

3.3 Policy Development Toolkit

The PDT, as described by deliverable D5.2 [56], is a web application which will include a front-end, allowing end-users to create and evaluate policy models, and a back-end which will handle data, model storage and other functionalities needed to provide end-users with the desired experience.

Concerning the back end, data storage and analytics have been addressed in previous sections.[54] It is important to note, additionally, that information on platform end-users will typically be collected and stored in this back-end for a variety of purposes, from user authentication to user analytics. This information will be classifiable as personal data to the extent that it can be traced back to an individual user and thus the following concerns, as a minimum, will need to be addressed:

1. *Lawfulness*. An appropriate legal basis for each purpose for which personal data on platform users may be collected should be identified, under Art. 6 GDPR. All steps needed to properly implement the identified legal bases should be taken: this will depend on the legal bases selected, which in turn depends on the manner in which user personal data may be processed via the PDT.[55]

2. *Fairness*. Personal data on PDT users should not be used in a manner which may violate their reasonable expectations. Profiling PDT users or covertly sharing their data with third parties for marketing purposes may qualify as unfair processing of personal data, in particular if this is not transparently disclosed to users. There should also be mechanisms in place to ensure that users are able to exercise their rights in relation to any personal data of theirs which may be collected during the use of the platform.

3. *Transparency*. A specific privacy policy should be developed for the PDT, in order to provide written information to users as to how their personal data may be handled when using the PDT in a concise, transparent, intelligible and easily accessible form, using clear and plain language, meeting all of the requirements of Arts. 13 and 14 GDPR. To the extent that cookies or similar tracking technologies are to be used on the platform, this should be clearly disclosed to users in the privacy policy, as well as in a separate cookie banner presented to users upon accessing the PDT, offering users relevant choices as to their cookie preferences.

4. *Purpose limitation*. Personal data collected on PDT users should not be used for purposes which are incompatible with those declared in the privacy policy.

5. *Data minimisation*. Only the strict minimum amount of user personal data needed to properly provide the PDT services, to perform any supporting functions, or for any other purpose for which a legal basis can be identified, should be collected.

6. *Accuracy*. Users should be given the possibility to rectify any personal data they submit, or which is collected on them, in connection with use of the PDT. This is linked to the above described fairness principle.

7. *Storage limitation*. Personal data on PDT users may only be stored for the strict minimum amount of time needed to meet any of the purposes for which they were lawfully collected, after which the personal data

[54] See Sections 3.1 and 3.2 above.
[55] As this is further specified during the Project, this section will be updated accordingly.

should be irreversibly anonymised or deleted. This will require the identification of appropriate retention periods for any such personal data collected.

8. *Security*. Appropriate security measures to ensure the confidentiality, integrity, and availability of any collected and further stored personal data, as well as the resilience of systems used to collect and further store those data, must be implemented.

9. *Accountability*. Records of evidence showing compliance with these requirements must be kept.

Additionally, from the contractual perspective, it will be important to define terms and conditions for the use of the PDT, so as to properly regulate the service relationship established between PolicyCLOUD and the end-user or the organisation to which the end-user belongs. These terms and conditions would need to be accepted for the use of the PDT to be allowed. Matters which should be regulated include:

1. Description of services offered through the PDT.
2. Payment terms (if applicable).
3. Acceptable use of the PDT.
4. Intellectual property (in particular, ownership of the PDT assets and the PDT output).
5. Data protection (with reference to the privacy policy, as well as to a data processing agreement or similar arrangement which may be put in place between the end-user and PolicyCLOUD). [56]
6. Warranties and PolicyCLOUD liability related to provision of the PDT, in particular, considering the likelihood of statistical inaccuracies in output presented by the PDT and the use of such output to create public policies.
7. Warranties provided by the end-user, and, in particular, compliance with legal and ethical requirements around selection of data sources and use of the PDT for policy-making purposes.
8. Applicable law.
9. Modification of the terms and conditions.
10. Termination and effects of termination.

[56] When offering the PDT as a service, PolicyCLOUD may be acting, simultaneously, as a controller for some activities involving personal data and as a processor on behalf of the end-users. As such, entering a standard DPA under Art. 28 GDPR may not suffice to fully regulate the data processing relationship between PolicyCLOUD and its end-users, as this would only cover the processor activities of PolicyCLOUD. To comprehensively regulate this relationship, the arrangement will need to include obligations to govern as well the PolicyCLOUD's as a controller. This can be done through a data management agreement, which distinguishes between the different sets of processing activities performed and allocates different corresponding sets of obligations to each party, to ensure that it is clear to what extent each party will be responsible for complying with the different controller obligations laid out in the GDPR.

11. Other matters which may require contractual regulation, depending on the specific service to be provided.

The data visualization components of the PDT should, as mentioned in Section 3.2.1.4 above, ideally not actually rely on or disclose information about identifiable individuals, but instead rely on appropriately aggregated data, which cannot be traced back to any given specific and identified individual, to avoid raising personal data protection concerns regarding data manipulated by PDT users. Where this is feasible, the core concerns around data visualization will be focused on:

1. *Statistical accuracy.* The same considerations as developed for analytics functions apply.[57]

2. *Explainability of results.* In order for an end-user to be able to use the PDT in a meaningful way, as a tool to support policy-making while preserving their own accountability for the decisions taken regarding any specific policy, it is important that end-users are able to understand the output presented to them by the PDT. Several different means can be conceived of to provide this information to end-users. The most appropriate form of delivery, considering the user experience and context when operating the PDT, should be further assessed. Clear explanations should be given to end-users on:

    - Steps taken across the design and implementation of the underlying platform to ensure compliance with legal and ethical requirements, to maximise the security and robustness of the platform, as well as the accuracy and reliability of the output.
    - The types of data sources and data used to produce outputs.
    - The rationale behind outputs, in terms of explaining generically how the data sources and data were processed to generate the outputs in question.
    - The likelihood of statistical inaccuracies in the outputs and the importance of critical examination and validation of output implications for policymaking by the end-user.

3.4 Data Marketplace

As the definition of requirements for the Data Marketplace component of the platform is still ongoing, the specific concerns from the legal, ethical, and societal perspective which may be raised will be addressed in a future update to this deliverable.

[57] See Section 3.2.1 above.

4 Specific issues related to use case #1 (policies against radicalisation)

The first use case will develop a collaborative data-driven analysis for the validation of existing policies against radicalization based on a participatory review of data coming from social media and open datasets. In addition, it will provide useful insights and valuable information to policy makers at any level (local, regional, national and EU level) to update current policies and/or create new ones, while at the same time allow them to interact with other relevant stakeholders (i.e., LEAs, social services, schools, and civil society) during the creation and modelling of new policies and specific countermeasures, ranging from early detection methodologies to techniques and policies for the monitoring and management of domestic radicalization. [56]

The use case objectives are through PolicyCLOUD big data streaming and real-time big data platform to improve operational efficiency, transparency, and decision making. The PolicyCLOUD visualization technologies will enable policy makers to identify issues, trends, and policy effects and interactions. The PolicyCLOUD analytics technologies will enable to discover insights and find meaningful explanations about the effects of policies.

Regarding this use case, three scenarios have been identified:

1. Scenario A consists in visualizing a heatmap that shows the frequency of occurrence of radicalization incidents in the geographic proximity of a region. Data coming from the GTD will be used. The policy maker can select the area of his/her interest and consult the different incidents that have taken place in each period. The related goals are to validate existing policies and investigate if there is a need to update them or create new one based on the retrieved information.

2. Scenario B consists in visualizing a bar chart that shows the main actors (individuals or groups) involved in radicalization efforts. Data coming from the GTD will be used. The policy maker can select the individuals and/or groups active in the area of his/her interest and consult the different incidents that are linked to each of them. The related goals are:

    - To identify main actors (individuals or groups) involved in violent incidents or propaganda spreading through online and offline activities.

    - To validate existing policies and investigate if there is a need to adjust and/or update them or create new one based on the retrieved information.

3. Scenario C consists in visualizing a bar chart that shows the main trends linked to radicalization. Data coming from the social media will be used. The policy maker can select the keywords of his/her interest and consult the different information linked to them. The related goals are to validate existing policies and investigate if there is a need to adjust and/or update them or create new one based on the retrieved information.

Regarding this use case, the main ethical, legal, regulatory, and societal concerns to be addressed are:

1. The exception to general citizens and data protection rights based on law enforcement and antiterrorism legal framework.

2. The impact on fundamental rights, as recognized by the CFREU and the common constitutional traditions of the EU member States.

3. The application of fairness and accuracy principles to avoid cognitive biases with regards to the parameters used to detect radicalization trends.

4. Conflicts related to the application of the transparency principle in the context of data processing and possible limitation to the exercise of the rights of the Data Subjects.

5. Choice between the processing of aggregate or not aggregate data to achieve the defined purposes.

To identify and tackle all potential legal and ethical issues related to this use case, the following specific objectives shall be targeted:

1. To evaluate data protection issues and impacts on the use case regarding the processing of personal data pursuant to the GDPR and EU Directive 2016/680.

2. To evaluate the Cybersecurity Strategy for the EU and ENISA activities.

3. To define the cooperation between the key actors involved in the use case.

4. To carry out an impact assessment concerning the possible ethical and legal risks for the persons involved in the use case.

5. To evaluate the impacts of the use case on human rights with regard on how users make use of social media and web, also for non-criminal and legitimate reasons and activities and evaluation of measures that might be taken to prevent abuses.

6. To evaluate possible ethical issues that may arise in connection to potential misuse of the tools employed by the project made by governments affected by a high level of corruption or other potential issues.

Therefore, the following sections identify and analyse the use case requirements from different angles, including:

1. Legal requirements based on personal data protection and data ownership, with a focus on EU law.

2. Ethical requirements designed to ensure rights, freedoms, and societal compliance.

3. Technical requirements relating to platform scalability, efficiency, reliability, and security.

From this perspective, developing these legal, ethical and technical requirements will provide a common vision and shared understanding of underlying concepts throughout the entire use case, to support the architecture design and the subsequent work of the other WPs, and integrate an ethics and data protection by design and by default approach.

Moreover, since the identifiability and the identity of radicalized subjects could be a goal of the final product, based on the purposes of the actors involved, in the following paragraphs we will analyse:

1. EU rules on privacy and data protection, guaranteeing a lawful and accountable behaviour to comply with these norms, applying the accountability principle to give the Consortium and the other use case actors the responsibility and ability to demonstrate compliance with the GDPR and the EU Directive 2016/680.

2. Ethical issues as covered by EU fundamental rights and freedoms that are associated with the use case objectives. Ethical norms are derived by the CFREU, to develop a fair decision-making process in designing the use case tools. We will therefore analyse the fundamental rights of the human beings to identify how rights, freedoms, privacy, and data protection can be guaranteed also when fighting radicalization phenomena and terrorism threats.

By building an ethics and data protection by design approach, and combining ethical and legal issues with technical requirements, we will explore the balancing between non-discrimination, human dignity, public security and data protection, to create a set of technical and legal requirements guiding the use case development.

In the light of the above, in the next paragraphs we will primarily outline the relevant framework with regards to:

1. The ethical and legal principles defined by the CFREU, the ECHR and all other applicable international, EU and national legislation, to ensure end-user acceptance and ethical compliance.

2. Data protection legal obligations related to privacy risks (i.e. GDPR and EU Directive 2016/680 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).

3. Best practices and key features for the use case and its processes, considering the relevant privacy laws and the LEAs tasks in fighting terrorism, to balance the traditional clash between privacy and domestic security. [57]

4.1 Ethical issues

This section analyses ethical issues to identify, from a normative perspective, the central elements and requirements that present a necessary condition for building trustworthy tools, eliminating any risk that could have a negative impact on the rights and freedoms of the individuals.

In this sense, ethical requirements will be built on the ethical principles constituted in the CFREU, the ECHR and all other applicable international and EU legislations.

This approach will allow the Consortium to have a clear vision of the possible impacts on the rights of individuals, ensuring tools ethical standards through an ethics-by-design strategy.

Particularly, PolicyCLOUD will aim to ensure respect for people and human dignity, fair distribution of research benefits and burden and protecting the values, rights and interests of the subjects involved in the research. Indeed, adequate attention shall be dedicated to the fact that the research results have the potential to be misused because the technologies developed by the Consortium could have a severe negative impact on human right standards if they are misapplied.

To prevent any intentional or unintentional bias existing ex ante, prior to the design and development of the system, the values of the designer or the values of end-users should be guided by common principles to be embedded into the system.

4.1.1 CFREU and ECHR rights and principles

In the EU legal environment and in all recitals of EU norms, one of the main constraints is to balance security and fundamental rights. Online security can only be sound and effective if it is based on fundamental rights and freedoms and the rights of the individuals.

Article 52 CFREU states that "*Any limitation on the exercise of the rights and freedoms recognized by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognized by the Union or the need to protect the rights and freedoms of others*". The CFREU provides safeguards

for fundamental human rights which may be only interfered by legitimate law enforcement activities. To do so, three elements must be considered:

1. What precisely is the national law to be considered, analysing to what extent it was accessible and cognizable. The interference must have some basis in domestic law and be compatible with the rule of law. Also, the law must be adequately accessible and foreseeable, that is, formulated with sufficient precision to enable the individual to regulate his or her conduct. [57]
2. Furthermore, since only a legitimate need can limit the rights and freedoms of individuals, there should be an evaluation of proportionality of that restriction for the purpose set by the provision, considering whether this contrast was justifiable insomuch as necessary in a democratic society. To be in pursuit of a legitimate aim requires that an activity is carried out in pursuance of one of the aims set out in Art. 8, par. 2 ECHR. [57]
3. The criterion of necessity should not be confused with an arbitrary judgment on the usefulness of the restriction because the interference must always respond to an urgent social need, be commensurate with the objective, and have adequate and relevant reasons.[58]

In this sense, protecting fundamental rights, freedom of expression, personal data and privacy needs to cope with the security need, proportionating safety and human dignity, freedom, democracy, equality, the rule of law and the respect for fundamental rights. For cyberspace to remain open and free, the same norms, principles, and values that the EU upholds offline, should also apply online. Fundamental rights, democracy and the rule of law need to be protected in cyberspace while protecting against incidents, malicious activities, and misuse.[59]

As the first use case aims to develop advanced tools for fighting online radicalization phenomena, this security purpose must be balanced with the following ethical pillars, established by the CFREU and the ECHR:

1. Human dignity[60], which includes respect for private and family life[61], protection of personal data[62], freedom of expression and information[63] which has to be interpreted as the right to produce, publish, transmit and share data (active profile), but also to be able to be informed by those who prepare and transmit news of public interest and furthermore to be able to access that news (passive profile). Already if these three basic profiles are considered, such freedom is also founded on the right to research information and sources and on guarantees of pluralism.

[58] See The Sunday Times v. The United Kingdom, No. 6538/74, §42, ECHR 1979.
[59] See Joint communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions, Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace [2013] JOIN (2013) 1 final, p. 2.
[60] See Art. 1 CFREU.
[61] See Art. 7 CFREU and Art. 8 ECHR.
[62] See Art. 8 CFREU.
[63] See Art. 11 CFREU.

2. Equality and non-discrimination, which includes equality before the law[64] and prohibition of any discrimination based on any ground[65], which is linked to the freedom of thought, conscience and religion[66].

3. Presumption of innocence and right of defence[67], which guarantees that everyone who has been charged shall be presumed innocent until proved guilty according to law. Moreover, anyone who has been charged has the right of the defence.

4. Principles of legality and proportionality of criminal offences and penalties[68], that excludes the possibility to declare someone guilty of any criminal offence on account of any act or omission which did not constitute a criminal offence under national law or international law at the time when it was committed.

4.2 Personal data protection and privacy

The following list summarizes the core elements that must be considered when implementing the use case to respect data protection and privacy principles:

1. *Data minimization*. The amount of data collected must be restricted to the minimum possible. These data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.[69]

2. *Purpose limitation*. Data must be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Art. 89, par. 1 GDPR, be not incompatible with the initial purposes.[70]

3. *Data quality and accuracy*. Personal data must be "*accurate and, where necessary, kept up to date*".[71]

4. *Storage limitation*. Data must be kept in a form which permits identification of Data Subjects for no longer than is necessary for the purposes for which the personal data are processed. Personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Art. 89, par. 1 GDPR subject to implementation of the appropriate technical and organizational measures required by the GDPR to safeguard the rights and freedoms of the data subject.[72]

[64] See Art. 20 CFREU.
[65] See Art. 21 CFREU.
[66] See Art. 10 CFREU.
[67] See Art. 48 CFREU.
[68] See Art. 49 CFREU
[69] See Art. 5, par. 1, let. c) GDPR.
[70] See Art. 5, par. 1., let. b) GDPR.
[71] See Art. 5, par. 1, let. d) GDPR.
[72] See Art. 5, par. 1, let. e) GDPR).

5. *Integrity, confidentiality, and security*. Personal data must be processed in a manner that ensures appropriate security, including protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organizational measures.[73]

6. *Accountability*. The controller shall be responsible for and be able to demonstrate compliance with all the data protection principles provided for by Art. 5, par. 1 GDPR.[74]

7. *Lawfulness of data processing*.[75] Personal data may be processed only if the data subject has given his/her consent for one or more specific purposes or if processing is necessary:

    - For the performance of a contract to which the Data Subject is party.
    - For compliance with a legal obligation to which the controller is subject.
    - To protect the vital interests of the data subject or another natural person.
    - For the performance of a task carried out in the public interest.
    - The purposes of the legitimate interests pursued by the controller.

8. *Consent to the processing.* It should consist in a freely given, specific, informed, and unambiguous indication of the wishes of the data subject about the data processing. The data subject has the right to withdraw his/her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal.[76]

9. *Special categories of personal data (also called "sensitive personal data").* The processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation can be performed only under very specific circumstances.[77]

Considering that the use case main actors and end-users include LEAs, which also can be defined as data controllers, EU Directive 2016/680 shall be duly considered. This directive establishes a set of rules and principles to be applied 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. So, even if the GDPR constitutes the legal background for the data processing, there are some specific duties and norms for data processed by competent authorities as the LEAs.

As anticipated above, the use case aims to realize a flexible prototype system with different settings and its tools include the possibility to collect personal data of radicalized individuals, which must be processed in compliance with the GDPR and the EU Directive 2016/680 principles. Under this directive, the personal data of the individual must be processed lawfully, fairly, and only for a specific purpose, which must be always linked to the fight against

[73] See Art. 5, par. 1, let. f) GDPR).
[74] See Art. 5, par. 2 GDPR).
[75] See Art. 6 GDPR.
[76] See Arts. 4, par. 11 and 7 GDPR.
[77] See Art. 9 GDPR.

crime. The directive ensures that personal data processing across the EU complies with the principles of rule of law, proportionality, and necessity, with appropriate safeguards for individuals. It also ensures completely independent supervision by national data protection authorities and effective judicial remedies.

The following list summarizes the core elements that must be considered when defining tools development and Consortium activities in the context of the first use case, to respect, since the design of the tools, data protection and privacy obligations and principles related to the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, as stated in EU Directive 2016/680:

1. *Time-limits for storage and review*. Even if the data protection principles are the ones established by the GDPR, there is a special provision about appropriate time limits to be established for the erasure of personal data or for a periodic review of the need for the storage of personal data. The competent authority must adopt procedural measures which ensure that those time limits are observed.[78]

2. *Distinction between different categories of data subjects*. There are four categories of data subjects affected by the EU Directive 2016/680:

    - Persons regarding whom there are serious grounds for believing that they have committed or are about to commit a criminal offence.
    - Persons convicted of a criminal offence.
    - Victims of a criminal offence or persons regarding whom certain facts give rise to reasons for believing that he or she could be the victim of a criminal offence.
    - Other parties to a criminal offence, such as persons who might be called on to testify in investigations in connection with criminal offences or subsequent criminal proceedings, persons who can provide information on criminal offences, or contacts or associates of one of the persons referred to in first two points.79

3. *Lawfulness of processing*. A competent authority can process personal data only for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security and these purposes must be based on EU or member State law. It also means that personal data collected by competent authorities for the above-mentioned purposes shall not be processed for other purposes unless such processing is authorized by EU or member State law. Where personal data are processed for such other purposes, including for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, the GDPR shall apply.[80]

4. *Accountability*. Also in case of a processing carried out by a competent authority, the controller (i.e., the LEA) shall be responsible for, and be able to demonstrate compliance with all the data protection principles provided for by Art. 4 EU Directive 2016/680, by implementing appropriate technical and organizational measures.

[78] See Art. 5 EU Directive 2016/680.
[79] See Art. 6 EU Directive 2016/680.
[80] See Arts. 8 and 9 GDPR.

5. *Processing of special categories of personal data*. The general limitations set by the GDPR about sensitive data processing are substituted by a general authorization for LEAs, but only where strictly necessary and with appropriate safeguards for the rights and freedoms of the data subject. Moreover, to avoid any discretional judgement made by the LEA, the processing of personal data must be:

   - Based on EU or member State law.
   - Performed to protect the vital interests of the data subject or of another natural person.
   - Related to data which are manifestly made public by the data subject.

   This kind of setting is particularly relevant in case of monitoring of religion-based terrorism groups and in preventing attacks made by them.[81]

6. *Data protection by design and by default*. These two principles recall Art. 25 GDPR.[82]

7. *Designation of data processors*. Where processing is to be carried out on behalf of a controller, the controller must use only processors providing sufficient guarantees to implement appropriate technical and organizational measures. Processing by a processor shall be governed by a written contract or other legal act under EU or member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of Data Subjects and the obligations and rights of the controller.[83]

8. *Records of processing activities*. Each controller and processor shall maintain a written record of processing activities under its responsibility, including, where applicable, the use of profiling and an indication of the legal basis for the processing operation, including transfers, for which the personal data are intended.[84]

9. *Logging*. Under EU Directive 2016/680, the controller and processor must keep logs for at least the following processing operations in automated processing systems: collection, alteration, consultation, disclosure (including transfers), combination and erasure. The logs of consultation and disclosure shall make it possible to establish the justification, date and time of such operations and, as far as possible, the identification of the person who consulted or disclosed personal data, and the identity of the recipients of such personal data. The logs are planned only for verification of the lawfulness of processing, self-monitoring, ensuring the integrity and security of the personal data, and for criminal proceedings.[85]

By analysing the general elements of the EU Directive 2016/680, some of the most important provisions concern rights and obligations such as:

1. *Automated individual decision-making*. Unless authorized by EU or member State law to which the controller (i.e., the LEA) is subject and which provides appropriate safeguards for the rights and freedoms of the Data Subject, controller decision based solely on automated processing, including profiling, which

[81] See Art. 10 EU Directive 2016/680.
[82] See Art. 20 EU Directive 2016/680.
[83] See Art. 22 GDPR.
[84] See Art. 24 GDPR.
[85] See Art. 25 EU Directive 2016/680.

produces an adverse legal effect concerning the Data Subject or significantly affects him or her, is prohibited. The same approach is followed for decisions based on special categories of personal data, unless suitable measures to safeguard the rights and freedoms and legitimate interests of the Data Subject are in place, to avoid discrimination against natural persons on the basis of special categories of personal data.[86]

2. *Information to be made available or given to the data subject*. The controller must make available to the data subject at least the following information:

    - The identity and the contact details of the controller.
    - The contact details of the DPO, where applicable.
    - The purposes of the processing for which the personal data are intended.
    - The right to lodge a complaint with a supervisory authority and the contact details of the supervisory authority.
    - The existence of the right to request from the controller access to and rectification or erasure of personal data and restriction of processing of the personal data concerning the Data Subject.

Member States may adopt legislative measures delaying, restricting or omitting the provision of the information to the Data Subject related to the legal basis for the processing, the period for which the personal data will be stored, or, where that is not possible, the criteria used to determine that period; the categories of recipients of the personal data, including those located in third countries or international organizations; further information, in particular where the personal data are collected without the knowledge of the Data Subject. These legislative measures can be adopted to the extent that, and for as long as, such a measure constitutes a necessary and proportionate measure in a democratic society with due regard for the fundamental rights and the legitimate interests of the natural person concerned, to:

- Avoid obstructing official or legal inquiries, investigations, or procedures.
- Avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or the execution of criminal penalties.
- Protect public security.
- Protect national security.
- Protect the rights and freedoms of others.87

3. *Right of access by the data subject and its limitations*. In case of processing under EU Directive 2016/680, the right of access has quite the same characteristics of Art. 15, par. 1 GDPR. It means that every citizen in the EU has an equal right of access to their personal data and they always have the right to approach the police and criminal justice authorities directly and ask for access to their personal data. However, member

[86] See Art. 11 EU Directive 2016/680.
[87] See Art. 13 EU Directive 2016/680.

States may adopt legislative measures restricting, wholly or partially, the right of access to the extent that, and for as long as such a partial or complete restriction constitutes a necessary and proportionate measure in a democratic society with due regard for the fundamental rights and legitimate interests of the natural person concerned, to:

- Avoid obstructing official or legal inquiries, investigations, or procedures.
- Avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or the execution of criminal penalties.
- Protect public security.
- Protect national security.
- Protect the rights and freedoms of others.

In the above-mentioned cases, the controller must inform the Data Subject, in writing, of any refusal or restriction of access and of the reasons for the refusal or the restriction, underlining the possibility for the Data Subject to lodge a complaint with a supervisory authority or seeking a judicial remedy.[88]

4.3 Soft regulation and best practices for the use case and its processes

The use case development depends not just on legal obligations, but also on other important features that involve the processes and platform. In this sense, several aspects that need to be handled in the frame of the use case have been identified also considering the GDPR and technical basic conditions to guarantee system functionalities:

1. *Minors of age involvement limitation*. Children inclusion into the investigation tool activities must be avoided, excluding the collection of data related to minors, if possible, or storage after accidental collection.
2. *Ethics by design approach*. To protect citizens' rights and freedoms, the whole platform design shall embed ethical requirements in all its technical and organizational measures and procedures, which will be an integral part of the accountability of the system.
3. *DPIA and ethical assessment*. As a fundamental part of data protection and ethics-by-design approach, the Consortium will carry out an assessment of the impact of imagined processing operations on the protection of personal data and ethical values, to identify and reduce the privacy risks and the likeliness of rights and freedoms infringements.
4. *Accountability in data protection and ethics*. The Project is based on effective procedures to report, document, and explain the measures implemented to comply with data privacy law and ethical requirements, to ensure the compliance with data protection legislation, CFREU and ECHR. The Consortium will guarantee a lawful and accountable data processing throughout the duration of the project. As for their very nature the use case tools will have a flexible system, with different settings based also on the purposes of the analysis and the investigations, the establishment of a data protection and

[88] Arts 14 and 15 EU Directive 2016/680.

ethics management system will avoid the potential misuse of research results and final products, both from partners of the Consortium and from external malicious actors. Through the application of the accountability principle and by producing all the relative compliance documents, the Consortium will be responsible for and be able to demonstrate compliance with the GDPR and EU Directive 2016/680.

5. *Risk minimization*. The use case shall take all the security measures that are appropriate for minimizing the risk that personal data may be destroyed, lost, accessed without authorization, or processed unlawfully or by moving away from the purposes for which the data was collected.

6. *Technical features*. To combine the tools development with legal and ethical requirements, general properties of the system that concern its openness and availability, but also its compliance with legal and ethical obligations will be monitored for all the duration of the project. The Consortium acknowledges that the work that will be conducted within the project involves the development of technologies and the creation of information that could potentially have substantial direct impact on personal data of individuals. That is why the Consortium will ensure the opportunity to customize the system according to the stakeholder's requirements and will take care of the trust level to achieve for each deployment. Multiple security mechanisms and technologies will ensure protection from malicious abuses.

5 Specific issues related to use case #2 (intelligent policies for the food value chain)

With regards to the specific ethical, legal, regulatory, and societal issues related to the second use case, the main concerns are:

1. Addressing potential concerns on data source selection.
2. Evaluating data protection issues and impacts on the use case regarding the processing of personal data pursuant to the GDPR.

This use case aims at creating tools which may allow the Government of Aragon and other potential end-users to understand trends relating to the Denomination of Origin and improve policies in this specific field. Such purpose will be pursued through the exploitation of data from public databases and social media platforms.

The main objectives of this use case are:

1. Improve investments in agri-food promotion by the Government of Aragon.
2. Facilitate tools and access to new technologies for small and medium producers, tools based on open data, social media analysis, opinion mining.
3. Improve the distribution of products thanks to the tool created in the project, a tool that allows them to search and compare prices and their positioning of their products and their rivals.
4. Support for decision-making in the investment of the different geographical areas with market study elements based on AI.
5. Bring the agri-food industry closer to new technologies.
6. Support policy makers in the design and modelling of new policies and updating existing ones.
7. Create stable working groups between producers and policy makers allowing to improve both communication and the development of new tools.

The related use case scenarios are:

1. Scenario A, aiming at visualizing the sale price of wine on the different specialized websites, with automatic warning systems that avoid penalties for contracts with large distributors. The goal is to achieve control of distribution prices of both its own products and those of the competitors, allowing to improve commercial policy.
2. Scenario B, aiming at visualizing the negative and positive opinions on social networks of the different products analysed allowing an automatic and immediate response to the end user. The goal is to create an immediate communication with the end users, knowing their impressions, both positive and negative, that will allow to interact with the end customers more directly.
3. Scenario C, aiming aims at analysing the trends in the wine sector through the collection of data from specialized websites. This action will allow to the stakeholder to know the trends in each of the markets

that are of interest: therefore, by knowing the trends in the sector, the same stakeholders will be able to adjust their diffusion policies taking into account all possible parameters.

For Scenario B in particular, by processing publicly available information posted by users on social media platforms such as Facebook, Instagram, Twitter, etc., it will be possible for end-users exploiting the platform to:

1. Identify the behaviour of the competitors and their strategies, both in sales and positioning.
2. Realise the impact of the campaigns on consumers by creating a profile of the type of consumers.
3. Collect customers' opinions and sentiments reflected in their comments/posts. [49]

Also, this use case considers the possibility to exploit data which comes from social media influencers with the most representative users and those who have more followers, as well as from specialized websites, in order to determine if the campaigns launched by end-users are well addressed.

5.1 Compliance Assessment around Selection of Data Sources

As noted in Section 2.2 above, applicable legal, regulatory, ethical and/or societal requirements may restrict an end-user's leveraging of a data source which might otherwise be considered appropriate from a practical perspective (i.e., useful towards the goal which the PolicyCLOUD end-user wishes to achieve).

For this use case, the following data sources have been indicated as relevant [75]:

1. Common Agricultural Policy (Dataset 9), described as being composed of "Semi-structured [information,] [v]irtuoso (triplets), JSON or XML", seemingly collected from the Aragón Open Data platform. [59]
2. Wine Register (Dataset 10), described as "SIGPAC reference, variety, cultivation year, area (Hec)", seemingly collected from the Aragón SIGPAC platform. [60]
3. Production data per grape variety (Dataset 11), described as "Production data per grape variety", without clarity as to where such data will be sourced from.
4. Wine varieties and brands information from Twitter (Dataset 12), described as "Relevant Twitter posts published by users about wine varieties, brands", collected from Twitter.
5. Wine varieties and brands information from Facebook (Dataset 13), described as "Relevant Facebook posts published by users about wine varieties, brands", collected from Facebook.
6. Wine varieties and brands information from Instagram (Dataset 14), described as "Relevant Instagram posts published by users about wine varieties, brands", collected from Instagram.
7. Wine varieties and brands information from LinkedIn (Dataset 15), described as "Relevant LinkedIn posts published by users about wine varieties, brands", collected from LinkedIn.
8. Wine varieties and brands information from the web (Dataset 16), described as "Relevant web news, trends published by users and experts about wine varieties, brands", seemingly collected from a variety of unspecified websites.
9. Wine varieties and brands information from the e-commerce (Dataset 17), described as "[Data sourced from] Relevant ecommerce websites about wine varieties, brands", seemingly collected from a variety of unspecified e-commerce websites.

The end-user must carry out a comprehensive assessment, following the guidelines noted in Section 3.1.2.1 above, to ensure that these data sources can be adequately leveraged, from a legal, ethical, and societal perspective. In particular, the following key points should be considered, for each data source:

Dataset 9

1. Given that the end-user is presumably responsible for management of this database, there should in principle be no contractual restrictions towards leveraging information obtained from the Aragón Open Data Platform for the purposes of the end-user. If this is not the case, then the end-user must assess whether the contractual terms applicable to the platform might prevent this – in particular, the Terms of Use applicable to the platform [61] – and obtain an appropriate authorisation from the relevant rights holder if needed.

2. On this point, the Terms of Use subject all reuse of data available on the platform to a Creative Commons Attribution 4.0 International license ("CC BY 4.0 License") [62], which allows free sharing[89] and adaptation[90] of those data, provided that attribution[91] (including the date of last update of those data) is given. However, the Terms of Use restrict users from altering the contents of those data or adapting those data in a way that distorts their meaning. This broad license suggests that the end-user is generally authorised to use Aragón Open Data as intended by the use case, from the contractual/intellectual property/database rights perspective.

3. While this is not clear, it is possible that leveraging this data source may imply the processing of personal data. The obligations around use of this data source, in case any personal data are processed, are further developed in Section 5.2 below.

4. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the population of the borough), considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

[89] Defined by the CC BY 4.0 License as the ability to "copy and redistribute the material in any medium or format".
[90] Defined by the CC BY 4.0 License as the ability to "remix, transform, and build upon the material for any purpose, even commercially".
[91] Defined by the CC BY 4.0 License as the obligation to "give appropriate credit, provide a link to the license, and indicate if changes were made", in "any reasonable manner, but not in any way that suggests the licensor endorses [the subsequent use made]". "Appropriate credit" means providing the name of the creator and attribution parties, a copyright notice, a license notice, a disclaimer notice, and a link to the material, if these are all supplied by the licensor. Any modifications to the underlying material must be indicated, as well as any previous modifications.

Dataset 10

1. Given that the end-user is presumably responsible for management of this database, there should in principle be no contractual restrictions towards leveraging information obtained from the Aragón SIGPAC Platform for the purposes of the end-user. If this is not the case, then the end-user must assess whether the contractual terms applicable to the platform might prevent this and obtain an appropriate authorisation from the relevant rights holder if needed.

2. While this is not clear, it is possible that leveraging this data source may imply the processing of personal data. The obligations around use of this data source, in case any personal data are processed, are further developed in Section 5.2 below.

3. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the population of the borough), considering the purpose for which the data source is to be used. The end-user should identify and document specific steps taken to address reliability and accuracy concerns detected.

Dataset 11

As little information has been provided on this data source, it is not clear which guidelines can be provided for its compliance assessment yet. However, the following terms should be considered:

1. Where the end-user is not responsible for the management of this data source, the end-user must assess whether any contractual terms applicable to the contact centre exist which might prevent the end-user from relying on this data source (and the data within) for its intended purpose, and obtain an appropriate authorisation from the relevant rights holder if needed.

2. Similarly, where the end-user is not responsible for management of this data source, the end-user should assess whether the data source may qualify as a protected database under the Database Directive (under copyright or sui generis protection), as implemented in the local laws applicable to the end-user, or whether any relevant parts of the data source may qualify for copyright protection under the local laws applicable to the end-user, and obtain an appropriate authorisation from the relevant rights holder if needed.

3. While this is not clear, it is possible that leveraging this data source may imply the processing of personal data. The obligations around use of this data source, in case any personal data are processed, are further developed in Section 5.2 below.

4. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the population of the municipality), considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

Datasets 12 to 15

1. The end-user should carefully assess the terms and conditions offered by the respective social media providers (Twitter, Facebook, LinkedIn), to ensure that data collected by these platforms can be leveraged for the purposes of the end-user. In particular, a thorough analysis of the general Terms and Conditions and specific terms offered by each platform to the end-user (whether through general terms of use, or more specific terms of use aimed at professional users, potentially related to platform APIs) is needed to ensure that the end-user can leverage these data as intended, without breaching any contractual obligations.

2. In particular, it should be noted that Twitter's terms of service [63] specify that a third-party is not allowed to access or search or attempt to access or search the services provided by Twitter by any means (automated or otherwise) other than through the interfaces provided by Twitter, unless a separate agreement has been entered with Twitter. Also, it is specified that crawling Twitter services is permissible if done in accordance with the provisions of the robots.txt file92, while scraping the services without the prior consent of Twitter is expressly prohibited. Similarly, Facebook's terms of service [64] prohibit to access or collect data on Facebook using automated means without prior Facebook permission and, provided that an agreement with Facebook is entered, it is nevertheless necessary to comply with specific terms governing the automated data collection. [65] LinkedIn's User Agreement [66] also explicitly states that users may not "Develop, support or use software, devices, scripts, robots or any other means or processes (including crawlers, browser plugins and add-ons or any other technology) to scrape the Services or otherwise copy profiles and other data from the Services", which suggests the need for a separate agreement or authorisation from LinkedIn in order to be able to do so.

3. Similarly, the end-user should assess whether these data sources may qualify as protected databases under the Database Directive (under copyright or sui generis protection), as implemented in the local laws applicable to the end-user, or whether any relevant parts of the data sources may qualify for copyright protection under the local laws applicable to the end-user, and obtain an appropriate authorisation from the relevant rights holder (which may the social media provider or the individual social media user) if needed.

4. It is strongly implied that leveraging this data source will involve the processing of personal data (this will particularly be the case whenever tweets, posts or other social media content are collected in a form which allows their poster/uploader/sharer to be identified, whether through their name, online handle, or other potential identifiers). The obligations around use of these data sources, in case any personal data are processed, are further developed in Section 5.2 below.

5. The end-user should identify steps to reasonably assure itself of the reliability of these data sources, in terms of the likelihood that any data collected from these data sources may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the target population), considering the purpose for which these data sources are to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

[92] Robots.txt is a file used by websites to let "bots" know if or how the website should be scrapped or crawled and indexed.

Datasets 16 and 17

1. The end-user should carefully assess the terms and conditions applicable to each website / e-commerce platform selected, to ensure that data available on these websites / platforms can be leveraged for the purposes of the end-user. In particular, a thorough analysis of the general Terms and Conditions and specific terms offered by each website / platform to the end-user is needed to ensure that the end-user can leverage these data as intended, without breaching any contractual obligations.

2. Similarly, the end-user should assess whether these data sources may qualify as protected databases under the Database Directive (under copyright or sui generis protection), as implemented in the local laws applicable to the end-user, or whether any relevant parts of the data sources may qualify for copyright protection under the local laws applicable to the end-user, and obtain an appropriate authorisation from the relevant rights holder (which may the social media provider or the individual social media user) if needed. This assessment will arguably need to be carried out per website / platform individually.

3. While this is not clear, it is possible that leveraging these data sources may imply the processing of personal data. The obligations around use of these data sources, in case any personal data are processed, are further developed in Section 5.2 below.

4. The end-user should identify steps to reasonably assure itself of the reliability of these data sources, in terms of the likelihood that any data collected from these data sources may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the target population), considering the purpose for which these data sources are to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

5.2 Specific concerns on privacy and data protection

In this Section, we will present an overview of the main issues which the end-user – acting as a controller – must bear in mind in the definition of requirements for use of the platform in each of the presented scenarios, with reference to the GDPR's data protection principles described in Section 2.2.4 above. Only those principles which present specific additional concerns to those provided in the general sections will be addressed – if a specific principle is not covered for this use case, the general section on that principle applies.

Of the different data sources indicated for this use case [75], it seems that the data sources identified as #12 to #15 (Wine varieties and brands information from Twitter / Facebook / Instagram / LinkedIn) present a clearer potential for capturing personal data than the remaining data sources. As such, this section will, in principle, be more relevant to those specific data sources (but should also be considered for the other data sources, in the event that use of any of the other data sources also implies a collection of personal data).

Getting now to examine the specific data protection implications underlying this use case, it shall first of all be noted that this specific use case [49] may entail the collection and processing of a different set of personal data, such as:

1. Data Subjects' names/surnames/ pseudonyms.

2. Personal opinions posted on social media platforms and – more generally – information publicly disclosed on social media platforms.

3. Profile of the relevant data subjects inferred from the automatic processing of personal data mentioned above.

For the purposes of this use case, the end-user shall be considered as an (independent) controller within the meaning of the GDPR, as it will use personal data available through the features of the platform for their own specific purposes. On the other hand, PolicyCLOUD will arguably act as a processor on behalf of the end-user, as in this phase it will only limit itself to making personal data available to the end-user on the platform, and further processing those personal data in accordance with the end-user's specifications, for the purpose defined by the end-user.

5.2.1 Lawfulness

As noted in Section 2.2.4.1 above, any use of personal data must be performed on the basis of consent provided by the Data Subjects, or otherwise on some other legitimate basis laid down in law, as set out in the GDPR or in other Union or Member State laws referred to by the GDPR.

Indeed, although personal data processed for this use case may be publicly available (e.g., on social media platforms, websites, blogs, etc.), this does not exempt the end-user from the obligation to find a suitable legal basis for the processing of such data for its own specific purposes. The end-user for this particular use case must therefore assess which of the legal bases afforded by the GDPR may be applicable and implementable. This assessment must consider the full context of the processing activities which are intended, including the specific data sources to be used (e.g., the types of personal data included in those data sources, the manner in which those data are collected – such as whether data are collected directly from data subjects, or indirectly from other data sources) and the specific goals to be reached through use of the platform.

To this end, considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.1 above can be provided on the particularities of specific legal bases which may potentially be relevant for consideration:

5.2.1.1 Consent

Consent will only be a feasible legal basis for the processing of personal data on social media users where all the requirements laid out in Section 2.2.4.1 above can be met in practice. In particular, given the potential inherent imbalance of power between the end-user for this use case (a public authority) and a social media user (particularly if that user is a citizen under the end-user's jurisdiction), it must be made absolutely clear to social media users that they will suffer no adverse consequences if they choose not to provide their consent, or to later withdraw it.

Furthermore, reliance on consent implies the need to ensure that consent given can be withdrawn, at which point all processing of personal data related to the consenting social media user must cease[93] and, in the absence of a legal basis to further process those data, deleted[94] – as such, consent can only be used where it is feasible to allow for the personal data of one or more social media users to be deleted/removed from further processing, should their consent be withdrawn[95].

[93] Art. 7, par. 3 GDPR.
[94] Art. 17, par. 1, let. b) GDPR.
[95] On this point, as noted by Art. 11, par. 2 GDPR, a controller is not required to comply with a data subject's request for the exercise of rights under the GDPR if it is no longer in a position to identify that data subject, except where

In this particular context, consent of data subjects may be tricky to obtain, as the very nature of the collection and processing of personal data makes it difficult, if not impossible, for the end-user to interact with each specific data subject posting contents which are considered relevant by the algorithms developed by PolicyCLOUD on social media or other public sources before the processing is carried out, in order to request their consent for the processing of their personal data. As such, other alternatives may need to be explored.

5.2.1.2 LEGITIMATE INTERESTS

As mentioned in Section 2.2.4.1 above, it is key to note that Art. 6, par. 1, second subparagraph GDPR does not allow public authorities, in the performance of their tasks, to rely on this legal basis. As such, if the end-user is acting as a public authority, in the performance of tasks mandated to it by law or regulation (as opposed to acting in the capacity of a private entity), this legal basis cannot be relied on by the end-user. To the extent that this is the case – which is not entirely clear for this use case – this legal basis is not available to the end-user for this use case.

Where Art. 6, par. 1, let. f) GDPR is available, as mentioned, the end-user will need to perform a balancing test / legitimate interest assessment, following the practical steps described in Section 2.2.4.1 above. This can be carried out as part of a broader data protection impact assessment (see Section 8.1.1 below). Where the interests of the end-user clearly outweigh the impact upon data subjects or can otherwise be supported by additional safeguards to mitigate the impact upon data subjects to an acceptable degree, this legal basis can be relied on.

5.2.1.3 PUBLIC INTEREST

As mentioned in Section 2.2.4.1 above, where it is not feasible for the end-user to rely on either of the above options, the end-user may consider whether it can justify the intended processing of personal data on the need to perform a task carried out in the public interest, or in the exercise of official authority[96]. As this legal basis presents the most flexible approach available to public authorities, when acting in the performance of their tasks, the end-user should assess whether the requirements described in Section 2.2.4.1 above for this legal basis are met, so as to ensure the lawfulness of the intended processing activities.

---

that data subject provides additional information enabling their identification. This may be relevant in a scenario where a citizen's personal data is promptly aggregated with that of other citizens, such that information pertaining to them is no longer linkable to them individually – in this case, it is reasonable to maintain that a withdrawal of consent would not prevent the controller from continuing to rely on the aggregated information for analytics purposes, insofar as it is not possible to establish a direct link between the aggregated information and the specific data subject. This would not apply, however, to any "raw" personal data kept on that data subject, which would be fully subject to their right to withdrawal of consent.

[96] Art. 6, par. 1, let. e) GDPR.

5.2.2 Lawfulness (Special Categories of Personal Data)

As noted in Section 2.2.4.2 above, where any special categories of personal data are to be collected and further processed, an applicable derogation to the GDPR's general prohibition on the processing of these personal data[97], from those listed in Art. 9, par. 2 GDPR, or as may be further provided under applicable Member State law[98], must also be identified. For clarity, to lawfully process special categories of personal data, the end-user must identify an applicable legal basis under Art. 6 GDPR AND an applicable derogation under Art. 9 GDPR.

To this end, considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.2 above can be provided on the particularities of specific derogations which may potentially be relevant for consideration:

5.2.2.1 Explicit consent

On this point, we refer to the requirements set out in Section 2.2.4.1 above, as all of those must be met to ensure validity of consent, and on the general observations as to the viability of reliance on consent for this use case made in Section 5.2.1.1 above.

5.2.2.2 Data manifestly made public by the data subject

As noted in Section 2.2.4.2 above, where a data subject has manifestly made personal data public, this may serve as a derogation to the general prohibition on use of special categories of personal data related to them[99]. This is a particularly relevant derogation for this use case, to the extent that personal data is to be collected from publicly available content uploaded on social media platforms. However, in this regard, the factors indicated by the European Data Protection Board, in the context of social media platforms, for determining whether this derogation is met – as described in Section 2.2.4.2 above – are particularly relevant. The following points should be borne in mind:

1. If social media platform's default settings, for the publication of content, are private, such that a data subject needs to actively choose to share data publicly, this weighs in favour of this derogation. As such, only data uploaded by social media users publicly (as opposed to data uploaded onto private profiles, groups, or shared through private direct messages, for example) should be considered.

2. Where a social media platform is meant for the public disclosure of data in a non-intimate or personal setting, such as is the case with LinkedIn (and arguably also Twitter), this weighs in favour of this derogation. A more careful assessment around Facebook and Instagram will be required, as Facebook and Instagram users may be less likely to expect a mass collection of content which they upload for the purposes pursued by the end-user.

3. If social media platforms inform their users that their content may be collected by other organisations, for purposes identical or similar to those pursued by the end-user, this weighs in favour of this derogation. For example, Twitter's Privacy Policy [67] informs users that "Twitter is public and Tweets are immediately viewable and searchable by anyone around the world", and that they may "share or disclose

[97] Established by Art. 9, par. 1 GDPR.
[98] In particular, regarding the processing of genetic data, biometric data or data concerning health, as set out in Art. 9, par. 4 GDPR.
[99] Art. 9, par. 2, let. e) GDPR.

non-personal data, such as aggregated information like the total number of times people engaged with a Tweet, demographics, the number of people who clicked on a particular link or voted on a poll in a Tweet (even if only one did), the topics that people are Tweeting about in a particular location, some inferred interests, or reports to advertisers about how many people saw or clicked on their ads". The end-user should make its own privacy policy publicly-available and easily accessible on its website(s) – as recommended in Section 2.2.4.4, above – to further support the likelihood of social media users becoming aware of how their data may be used.

The end-user should carefully assess whether this derogation may be applicable to any special categories of personal data (if any are collected) extracted from Twitter, Facebook, Instagram, LinkedIn or any other data sources it wishes to rely on.

5.2.2.3 SUBSTANTIAL PUBLIC INTEREST

As noted in Section 2.2.4.2 above, this derogation requires the demonstrable need to process special categories of personal data in order to meet a substantial public interest with a basis in Union or Member State law applicable to the controller which must (1) be proportionate to the interest pursued, (2) respect the essence of the right to data protection, and (3) provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests.

The end-user should assess whether this derogation may be applicable, considering the purposes which it may be seeking to pursue through processing of personal data via the platform.

5.2.2.4 STATISTICAL PURPOSES

In light of the requirements explained in Section 2.2.4.2 above, as the processing of special categories of personal data, if relevant, may be carried out in this use case for statistical purposes, the end-user should determine whether this processing can be based on Union or Member State law applicable to the end-user, which must (1) be proportionate to the interest pursued, (2) respect the essence of the right to data protection, and (3) provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests, this derogation may be applicable.

In this case, the end-user and PolicyCLOUD must collaborate to ensure that the specific safeguards further specified in Art. 89, par. 1 GDPR, as described in Section 2.2.4.2 above, can be implemented.

5.2.3 Fairness

As noted in Section 2.2.4.3 above, for the processing to be considered as fair under the GDPR, the end-user shall ensure that personal data are handled in ways that may be reasonably expected by data subjects and not use such data in a way that may produce unjustified adverse effects on them.

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.3 above can be provided on specific key elements to be considered by the end-user, under the principle of fairness:

5.2.3.1 Interaction

Given that the end-user will be collecting data indirectly from social media platforms, this requires the end-user to have channels in place through which data subjects can communicate with them, including to exercise their rights as data subjects (under the GDPR and local laws). As noted under the principle of transparency, addressed in Section 2.2.4.4 above, this also requires the end-user to take reasonable steps towards ensuring that data subjects are aware that their personal data may be processed in the manner intended by this use case – such as by publishing information on these activities on a publicly available website managed by the end-user.

5.2.3.2 Expectation

This element is particularly important in this use case, as depending on what information about the data subject are made available to the end-user through the platform and the consequent actions the end-user may perform, the processing of personal data may be considered as unfair (due to being reasonably unexpected) pursuant to the GDPR. In particular, while it may be argued that the collection of data subjects' publicly available comments/posts/interactions on social media platforms and websites and their subsequent aggregation in order to provide relevant statistics, trends, charts, etc. to shape product strategies may, to some extent, be expected by data subjects when they publicly post or comment on social media platforms (or, at least, could arguably be seen as not excessively intrusive), the same could not be stated should the platform allow the storage of data subjects' opinions on social media in such a way that allows the end-user to directly identify data subjects, trace content back to the original source (e.g. through a hyperlink) and interact with those data subjects. This level of intrusiveness may potentially be considered as excessive and, as such, run afoul of this principle. It should further be noted that such an activity could also be considered as out-of-scope in the context of the policy-making purposes for which the PolicyCLOUD platform is being developed, appearing more as a form of online brand protection/management. As such, it is strongly recommended that the end-user consider, whenever possible, the aggregation of personal data collected from public sources, to mitigate the risk of unfair processing.

One possible exception which could be borne in mind would be the case of processing unaggregated personal data on social media influencers. The inherent role played by influencers (which involves a greater deal of public exposure) may lead to suggest that they have a lessened expectation of privacy regarding content which they make publicly available on social networks, such that the processing of such content by a third party for the purposes pursued by the end-user would not be an unreasonable violation of their expectations. This should be carefully assessed by the end-user when identifying the legal basis applicable to them (e.g., in the legitimate interests assessment performed, where feasible, or in the assessment as to whether such an activity may be considered as performed in pursuit of a task in the public interest).

5.2.3.3 NON-DISCRIMINATION

Personal data should not be collected on social media users for the purpose of discriminating against them (such as to cause harm or detriment to social media users publishing content seen as problematic by the end-user), nor should this be the end-result of policies developed using social media users' personal data.

5.2.4 Transparency

To ensure its compliance with the principle of transparency, as seen in 1.2.4.4 above, the end-user must ensure that it provides complete and understandable information to data subjects on their data processing practices.

Ideally, this would involve the development of an information notice, to be provided directly to social media users upon collection of their personal data, in writing. However, given that personal data is not collected directly from data subjects in this use case (but rather from, e.g., social media platforms), the end-user may be able to argue for the exemption under Art. 14, par. 5, let. b) GDPR, where the provision of information directly to each individual social media user this proves impossible [39], or would represent a disproportionate effort for the end-user. [39] The end-user should develop a specific assessment to demonstrate that this exception is applicable, in particular contrasting the effort required of the end-user to ensure that each individual social media user would receive information directly against the harm which may arise for social media users should they not have any access to such information (e.g., should they not become aware of the processing carried out by the end-user). This can be carried out as part of a broader data protection impact assessment (see Section 8.1.1 below). Where the effort required of the end-user would be clearly disproportionate considering the (low) impact upon data subjects, this exemption can be relied on.

In this case, the end-user must take appropriate measures to ensure the protection of the rights and freedoms of social media users regardless of the fact that this information is not directly provided to them, such as by displaying the information on a publicly-available website, as stated in Art. 14, par. 5, let. b) GDPR. Any information notice or privacy policy developed (e.g., to make available on a website managed by the end-user, as seen above) must bear in mind the inherent tension between the GDPR requirements of providing comprehensive information, and ensuring that the information provided is concise, transparent, intelligible and easily accessible in mind – this requires an assessment as to which information should be prioritised, what the appropriate level of detail is and which are the best means by which to convey this information to data subjects. [39] Whenever feasible, the end-user should rely on the so-called "layered approach", allowing them to structure the information into relevant categories which the data subject can select, to ensure immediate access to the information deemed most relevant by the data subject and prevent information fatigue. [39] [68]

Whenever personal data can lawfully be processed by the end-user for a further purpose (i.e., where this further purpose is compatible with the original, or where an additional legal basis exists for the further purpose), the information made available on the end-user's website must be promptly updated, under Art. 14, par. 4 GDPR.

5.2.5 Purpose Limitation

As seen in Section 2.2.4.5 above, to ensure compliance with the principle of purpose limitation, under Art. 5, par. 1, let. b) GDPR, the end-user must identify specific, explicit, and legitimate purposes for which personal data are to be collected and processed, and then refrain from using personal data for any other incompatible purpose. Through the presumption established for "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" in Art. 5, par. 1, let. b) GDPR, provided that the safeguards of Art. 89, par. 1 GDPR are respected – as addressed above, in Section 2.2.4.2 – it is arguably possible for the end-user to make further use of personal data collected in a commercial context (such as in the context of social media platforms) for statistical analysis aimed at the pursuit of a public interest.

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.5 above can be provided on specific key elements to be considered by the end-user, under the principle of purpose limitation:

5.2.5.1 Specificity

The specific purposes for which the end-user intends to process data collected from social media platforms should be made clear in the end-user's publicly available information notice, as mentioned above, in Section 5.2.4.

5.2.5.2 Necessity

The end-user should, in particular, carefully assess whether the purposes for which it intends to use the platform can be used with only anonymous or aggregated social media data (preferred approach), as opposed to identifiable data (e.g., posts or content linked to a specific social media user through a name or social media handle, or through other identifiers).

5.2.5.3 Compatibility

The presumption of compatibility mentioned above arguably applies here, for research and statistical purposes. However, taking into account the general scope of the PolicyCLOUD Project and the purpose for which the cloud-based platform for this use case is developed, the end-user should generally avoid processing personal data to track or target specific social media users and directly interact with them on social media platforms, as this purpose may arguably be considered as excessive and unlawful (thereby incompatible with the other purposes for which the end-user may wish to use those data), as specified under Section 5.2.3.2 above.

5.2.5.4 Limitations to further processing

The point raised above on tracking or targeting of specific social media users is relevant for this element also.

5.2.6 Data minimisation

As seen in Section 2.2.4.6 above, compliance with the principle of data minimisation requires a minimalistic approach to personal data, in the sense that (1) as little of it as possible should be processed in order to meet an intended purpose, and (2) only personal data which are adequate, relevant and strictly necessary to meet a purpose should be used. Ultimately, if a purpose can be met without using personal data (e.g., using only anonymous or aggregated data), then no personal data should be used at all.

Considering the data sources identified by the relevant end-user for this use case [75], aside from the general guidance provided in Section 2.2.4.6 above under the principle of data minimisation, the end-user should arguably assess, for example, whether there is any added value in retaining the ability to identify the specific social media user uploading content onto a social media platform for the policy-making purposes which are pursued. If so, the end-user should then consider whether preserving this added value serves a legitimate goal, if the added value is substantial, and if the benefits of retaining this ability outweigh the potential impact on the social media users in question. If it determines that the use of personal data, preserving a link to the identity of individual social media users, is necessary, then the end-user must be able to demonstrate that each data point collected is specifically relevant to the purpose pursued. Any irrelevant personal data will be deemed as excessive and should not be collected or further processed.

5.2.7 Accuracy

Ensuring accuracy of data used is fundamental from the legal perspective – as seen in Section 2.2.4.7 above – but also from the ethical perspective, given that inaccurate, incomplete, misleading or biased data can result in erroneous outputs, culminating in misguided policy-making with a potential impact at an individual and societal level – as seen in Section 2.1 above.

Given that data will be extracted directly from social media platforms and websites, the principle of accuracy will generally be met in the sense that it should be easy to objectively demonstrate that a given social media user / individual factually uploaded a given piece of content; however, the more subjective analysis of ensuring the accuracy of data/information contained within uploaded content (and of the opinions / sentiments which can be derived from such content via the platform) is another matter.

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.7 above can be provided on specific key elements to be considered by the end-user, under the principle of accuracy:

5.2.7.1 Data source reliability

Where social media platforms are concerned, reliability of content uploaded will always be a relevant risk to be mitigated (as those platforms generally do not exercise any editorial powers over the content uploaded by social media users). The end-user, in collaboration with PolicyCLOUD, must identify measures to mitigate the risk of reliance on false or misleading data, to the extent that this may affect the outcome of the policy-making process.

5.2.7.2 Verification

Given that this use case will rely on several Streaming data sources, a continuous effort to ensure accuracy will need to be carried out, namely by overwriting older data with newer data, to prevent excessive data aging, through specific and adequate “rollover” periods to be determined by the end-user.

5.2.7.3 UPDATED DATA

As noted above, whenever a Streaming data source is used – which is the case for the social media platforms considered – appropriate rollover periods (i.e., periods after which older data is to be overwritten by newer data collected) should be defined.

5.2.8 Storage limitation

As noted in Section 2.2.4.8 above, this principle becomes practically relevant whenever personal data is collected, provided that it is not promptly anonymised or aggregated (with the underlying raw data being deleted as soon as possible). Noting that personal data avoidance is the preferred approach under the data minimisation principle, whenever this is not feasible, then the end-user (as controller) should define specific retention periods for the personal data collected, based on the strict minimum period of time for which retention of those data is needed to ensure that the purpose for their collection and processing can be met.

Considering the data sources identified by the relevant end-user for this use case [75], aside from the general guidance provided in Section 2.2.4.8 above, the end-user should consider that, when dealing with Streaming data sources, it is important for the end-user to define an appropriate retention/overwrite period to avoid data aging (i.e., defining a short "rollover" period after which older data will be overwritten by newer data collected through the stream). The platform should allow the end-user to define retention periods and include tools for automatic deletion or aggregation of underlying data after an end-user-defined retention period is exceeded.

5.3 Specific concerns related to ethical / societal impact

Aside from the ethical and societal concerns highlighted in Section 2.1, specific main concerns for this use case include:

1. The adoption of appropriate data protection and security measures when using the PolicyCLOUD platform.
2. The risk that reliance on platform output may be affected by a "filter bubble", leading to a failure to consider such output in the broader context in which the end-user operates (i.e., the autonomous community of Aragón), as suggested in Section 2.1.3 above.
3. The risk that policies built on platform output may favour certain territories and/or grape or other foodstuff varieties, without duly taking into consideration the relevant impact this may have on the autonomous community's biodiversity and natural ecosystems.

6 Specific issues related to use case #3 (urban policy making through analysis of crowdsourced data)

In this section we will analyse in greater detail the specific ethical, legal, regulatory, and societal issues related to the third use case. To this regard, the main concerns to be addressed are:

1. the assessment of the legal basis of processing with regards to data on criminal convictions and offences according with Art. 10 GDPR (e.g. with regards to the violation of public order). Indeed, the processing of this data shall be performed only under the control of official authorities and/or when the processing is authorised by EU or member States law providing for appropriate safeguards for the rights and freedoms of data subjects.

2. the assessment of the legal basis of processing with regards to special categories of data according with Art. 9 GDPR (for example, data included in specific authorizations in case of disabilities).

The aim of this use case is to support Sofia Municipality's policy making in important areas of citizen's areas of everyday life by using crowdsourced data via its contact centre. By improving the policy making in these areas, the overall quality of citizen's life will be improved, which is the overall goal of this project.

By using the powerful tools provided by PolicyCLOUD, Sofia Municipality will be able to carry out a detailed analysis of the territorial distribution of the signals by categories and types, areas, districts, major transport roads, etc. The results of the analysis will allow the municipal and district administrations to identify the problems in the urban environment and to adopt or modify adequate policy making decisions on budget planning and effective use of budget and public resources. It will also help Sofia Municipality to be focused on improving its policy making, related to better control and monitoring in these sectors, as well as preventing and avoiding risky or conflicting situations from happening.

The related use case scenarios are:

1. SC1 (Transport), aiming at improving the quality of service and transport times, achieving better connections for citizens and assess multimodal pricing schemes and initiatives such as the so-called "green ticket".

2. SC2 (Parking), which aims at adopting quantity measures for better parking management and improving overall parking capabilities.

3. SC3 (Road infrastructure), aiming at improving long term policy making in the area of road infrastructure and achieving better envisioning and capacity building of district administrations and municipal administration in solving road infrastructure problems.

4. SC4 (Waste collection and waste disposal), which aims at achieving a more efficient way of waste collection and the improvement of long-term planning and policy making of waste collection and waste disposal using smart meters.

5. SC5 (Air quality), which aims at the improvement of long-term policy making in the area of air quality.

The common goals of the above described scenarios are to validate existing policies and investigate if there is a need to update and/or modify them or create new one based on the retrieved information. [49][69]

6.1 Compliance Assessment around Selection of Data Sources

As noted in Section 2.2 above, applicable legal, regulatory, ethical and/or societal requirements may restrict an end-user's leveraging of a data source which might otherwise be considered appropriate from a practical perspective (i.e., useful towards the goal which the PolicyCLOUD end-user wishes to achieve).

For this use case, the following data sources have been indicated as relevant [75]:

1. Sofia Municipality Signals (Dataset 18), described as being composed of "Structured information, texts and images" collected from "Signals from citizens, coming through the contact centre of the municipality"; and

2. Sofia Municipality Airthings Platform (Dataset 19), described as "Structured information" collected from "Real time IoT sensors data for monitoring and measurement of the air quality".

The end-user must carry out a comprehensive assessment, following the guidelines noted in Section 3.1.2.1 above, to ensure that these data sources can be adequately leveraged, from a legal, ethical, and societal perspective. In particular, the following key points should be considered, for each data source:

Dataset 18

1. Given that the end-user is presumably responsible for management of the contact centre of the municipality, there should in principle be no contractual restrictions towards leveraging information obtained via the contact centre for the purposes of the end-user. If this is not the case, then the end-user must assess whether any contractual terms applicable to the contact centre exist which might prevent this and obtain an appropriate authorisation from the relevant rights holder if needed.

2. Similarly, where the end-user is not responsible for management of the contact centre of the municipality and for the compilation of information received through that contact centre, the end-user should assess whether the data source may qualify as a protected database under the Database Directive (under copyright or sui generis protection), as implemented in the local laws applicable to the end-user, or whether any relevant parts of the data source may qualify for copyright protection under the local laws applicable to the end-user, and obtain an appropriate authorisation from the relevant rights holder if needed.

3. While this is not clear, it is possible that leveraging this data source may imply the processing of personal data. The obligations around use of this data source, in case any personal data are processed, are further developed in Section 6.3 below.

4. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the population of the municipality), considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

Dataset 19

1. The end-user should carefully assess the terms and conditions offered by Airthings ASA (the provider of the Airthings platform), to ensure that data collected by Airthings can be leveraged for the purposes of the end-user. A thorough analysis of the general Terms and Conditions and specific terms offered by Airthings to the end-user is needed to ensure that the end-user can leverage Airthings data as intended, without breaching any contractual obligations.

2. On this point, the terms made available by Airthings online state that "Airthings grants the Customer a limited, nonexclusive, non-sublicensable, non-transferable license under intellectual property rights to use the Airthings application programming interface ("Airthings API") for the purpose of developing and implementing customer specific software solutions, products and applications integrating with products and services" (Section 4 of the Specific Terms for Subscription Services). This broad license suggests that the end-user is generally authorised to use Airthings data as intended by the use case.

3. While this is not clear, it seems that leveraging this data source will, in principle, not imply the processing of personal data (except for personal data on specific individuals authorised by the end-user to operate the Airthings API). [71] Specific concerns around the use of IoT technology are addressed in Section 6.2, below. To the extent that any personal data is processed through this data source, the obligations further developed in Section 6.3 below, while primarily aimed at Dataset 18, should also be considered for Dataset 19, with necessary adaptations.

4. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete, considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected. In this respect, specific provisions of the Airthings terms and conditions should be borne in mind, such as the fact that Airthings does not warrant the accuracy of data collected [70], the possibility for Airthings to apply rate-limiting to traffic or quotas to API usage, and the possibility for Airthings to introduce new API updates and versions at its discretion [70], all of which create risks to the accuracy and availability of data collected through the API.

6.2 Use of IoT technology

Regarding this use case, specific attention should be dedicated to the use of IoT technologies, from which stems the necessity for PolicyCLOUD and the policy makers to achieve end-user engagement and build trust among the citizens with regards to this technology. Indeed, during the last decade, the IoT has gained huge importance as it aims to provide people with innovative and intelligent technologies and services, in which all of the physical objects around them are linked to the internet and are able to communicate with each other. We have witnessed the evolution of the traditional internet into a global network of an enormous number of devices which are currently available to collect data that not only gather information from the physical environment but are designed to interact with people. Citizens have become sources of information and, at the same time, users of the information elaborated and provided by smart objects. Different security challenges could face the adoption of the IoT:

1. First, data anonymity, confidentiality and integrity are desirable to ensure the basic security concerns of end-users.

2. Moreover, access controls, which control authentication and authorization, is required to prevent unauthorized access to the system.

In this scenario, concerns with security and privacy regarding computer networks are always increasing. This kind of data processing has led to numerous discussions about the trade-off between the risks for the data protection of individuals and opportunities for the industry that arise from the analysis of such data sets. Those kinds of concerns have to be tackled in a trust-oriented approach, to fulfil the gap between the expectation in efficiency and the lack of information about the elaboration process and data usage that could lead citizens to lose interest and do not behave naturally in the interaction with IoT. The challenge is to develop technologies that are inherently privacy-preserving and may offer the basis for empowering the end-users and more in general the end-targets to understand and be informed of and, where appropriate, control over the use of their personal data, within the meaning of Art. 4, par. 1 GDPR.

Technology acceptance is a first step to beneficially use the IoT. Once accepted, the IoT potentially offers several benefits as it enables individuals to make better decisions. To adequately address the matter, we must consider not only end-users but also data subjects, whose data are being collected and processed through IoT even if not in an interactive usage. [72]

6.3 Specific concerns on privacy and data protection

In this Section, we will present an overview of the main issues which the end-user – acting as a controller – must bear in mind in the definition of requirements for use of the platform in each of the presented scenarios, with reference to the GDPR's data protection principles described in Section 2.2.4 above. Only those principles which present specific additional concerns to those provided in the general sections will be addressed – if a specific principle is not covered for this use case, the general section on that principle applies.

Of the two data sources indicated for this use case [75], it seems that the data source identified as #18 (Sofia Municipality Signals) presents a clearer potential for capturing personal data than the data source identified as #19 (Sofia Municipality Airthings Platform). As such, while the previous section was more relevant to the latter, this section will, in principle, be more relevant to the former.

6.3.1 Lawfulness

As noted in Section 2.2.4.1 above, any use of personal data must be performed on the basis of consent provided by the individuals whose data is used ("data subjects"), or otherwise on some other legitimate basis laid down in law, as set out in the GDPR or in other Union or Member State laws referred to by the GDPR.

To this end, considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.1 above can be provided on the particularities of specific legal bases which may potentially be relevant for consideration:

6.3.1.1 LEGAL OBLIGATION

As noted in Section 2.2.4.1 above, one legal basis which may potentially be applicable is the need to process personal data to comply with a legal obligation to which the end-user is subject[100]. It is recommended, as a first step in the lawfulness assessment, for the end-user to assess whether this legal basis may apply to the processing activities it may envision performing through PolicyCLOUD – i.e., if the requirements described in Section 2.2.4.1 above for this legal basis are met for this specific use case – before considering possible alternatives.

6.3.1.2 CONSENT

As noted in Section 2.2.4.1 above, where the above legal basis is unavailable, another which may apply is consent provided by data subjects for use of their personal data, for the purpose of generating aggregated information from which the end-user may be able to draw insights into relevant trends and issues arising at the city-level, so as to use those insights to focus policy-making on the most pressing issues and improve decision-making efficiency and effectiveness. This legal basis may be particularly relevant for the data source consisting in signals provided by Sofia citizens through the municipality's contact centre[101].

In order to obtain valid consent from citizens for this further processing of their personal data, the end-user should consider the feasibility of requesting consent from citizens when they decide to submit a signal, in a manner which meets all of the consent requirements of the GDPR (as described in Section 2.2.4.1 above). In particular, the end-user should bear in mind that making this consent mandatory in order to allow data subjects to benefit from the provision of a service (e.g., where a citizen would not be allowed to file a complaint with the municipality contact centre without consenting to further processing of their personal data for analytics purposes) would affect the freedom of their consent and, therefore, its validity. Furthermore, given the inherent imbalance of power between the end-user for this use case (a public authority) and an individual citizen, it must be made absolutely clear to citizens that they will suffer no adverse consequences if they choose not to provide their consent, or to later withdraw it – in particular, they must be clearly informed that their signal/complaint will still be duly processed by the municipality, even if they do not provide this additional consent.

6.3.1.3 PUBLIC INTEREST

As noted in Section 2.2.4.1 above, where neither of the above two options are available, the end-user may consider whether it can justify the intended processing of personal data on the need to perform a task carried out in the public interest, or in the exercise of official authority[102].

Given that it is not possible for a public authority to rely on the legal basis set out in Art. 6, par. 1, let. f) GDPR in the performance of their tasks, this legal basis presents the most flexible approach available to public authorities under those circumstances, and thus the end-user is recommended to assess whether, for this use case, the requirements for this specific legal basis – as described in in Section 2.2.4.1 above – are met.

[100] Art. 6, par. 1, let. c) GDPR.
[101] Dataset 18, as identified in **D1.3** [75].
[102] Art. 6, par. 1, let. e) GDPR.

6.3.2 Lawfulness (Special Categories of Personal Data)

As noted in Section 2.2.4.2 above, where any special categories of personal data are to be collected and further processed, an applicable derogation to the GDPR's general prohibition on the processing of these personal data[103], from those listed in Art. 9, par. 2 GDPR, or as may be further provided under applicable Member State law[104], must also be identified. For clarity, to lawfully process special categories of personal data, a controller must identify an applicable legal basis under Art. 6 GDPR AND an applicable derogation under Art. 9 GDPR.

To this end, considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.2 above can be provided on the particularities of specific derogations which may potentially be relevant for consideration:

6.3.2.1 Explicit consent

On this point, we refer to the requirements set out in Section 2.2.4.1 above, as all of those must be met to ensure validity of consent, and on the general observations as to the viability of reliance on consent for this use case made in Section 6.3.1.2 above.

6.3.2.2 Data manifestly made public by the data subject

As noted in Section 2.2.4.2 above, where a data subject has manifestly made personal data public, this may serve as a derogation to the general prohibition on use of special categories of personal data related to them[105]. To the extent that signals/complaints filed by citizens through the municipality's contact centre are made public, this derogation may potentially be considered. However, in this regard, the factors indicated by the European Data Protection Board for determining whether this derogation is met – as described in Section 2.2.4.2 above, in the context of social media platforms – are particularly relevant. The following additional points should be borne in mind:

1. If signals/complaints are, by default, kept confidential, such that a citizen needs to actively choose to share data publicly, this weighs in favour of this derogation. As such, this derogation would only potentially apply to those signals/complaints made public (and not to any kept confidential).

2. The end-user should carefully assess whether the context of the municipality's contact centre is such that an average citizen might reasonably expect that special categories of personal data which they disclose in public signals/complaints would be further used by the end-user for its intended purposes.

The end-user should carefully assess whether this derogation may be applicable to any special categories of personal data (if any are collected) extracted from signals/complaints filed by citizens.

[103] Established by Art. 9, par. 1 GDPR.
[104] In particular, regarding the processing of genetic data, biometric data or data concerning health, as set out in Art. 9, par. 4 GDPR.
[105] Art. 9, par. 2, let. e) GDPR.

6.3.2.4 SUBSTANTIAL PUBLIC INTEREST

As noted in Section 2.2.4.2 above, this derogation requires a basis in Union or Member State law applicable to the controller which must (1) be proportionate to the interest pursued, (2) respect the essence of the right to data protection, and (3) provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests.

The end-user should assess whether this derogation may be applicable, considering the purposes which it may be seeking to pursue through processing of personal data via the platform.

6.3.2.5 STATISTICAL PURPOSES

In light of the requirements explained in Section 2.2.4.2 above, as the processing of special categories of personal data, if relevant, may be carried out in this use case for statistical purposes, the end-user should determine whether this processing can be based on Union or Member State law applicable to the end-user, which must (1) be proportionate to the interest pursued, (2) respect the essence of the right to data protection, and (3) provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests, this derogation may be applicable.

In this case, the end-user and PolicyCLOUD must collaborate to ensure that the specific safeguards further specified in Art. 89, par. 1 GDPR, as described in Section 2.2.4.2 above, can be implemented.

6.3.3 Fairness

As noted in Section 2.2.4.3 above, for the processing to be considered as fair under the GDPR, the end-user shall ensure that personal data are handled in ways that may be reasonably expected by data subjects and not use such data in a way that may produce unjustified adverse effects on them.

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.3 above can be provided on specific key elements to be considered by the end-user, under the principle of fairness:

6.3.3.1 EXPECTATION

If data subjects are led to believe that personal data collected on them will be used to improve the municipality's policy-making abilities, this should be the only objective pursued with those personal data – using them to profile and target individuals raising problematic complaints, or for other unrelated and arguably illegitimate purposes (e.g., sending of marketing communications), must be strictly avoided. This will imply controls around purpose limitation, including access control.

6.3.3.2 NON-DISCRIMINATION

The controller shall not discriminate against data subjects. In particular, personal data should not be collected on citizens for the purpose of discriminating against them (such as to cause harm or detriment to citizens filing larger numbers of complaints), nor should this be the end-result of policies developed using citizens' personal data – this requirement is strongly tied to applicable ethical considerations of avoidance of bias and non-discrimination, as seen in Section 2.1.3 and 2.1.4 above.

6.3.3.4 NON-EXPLOITATION

The controller shall not exploit the needs or vulnerabilities of data subjects. Considering the inherent imbalance of power between the controller (a public authority) and individual citizens, this is particularly relevant when assessing the freedom of consent, where this legal basis is leveraged (as seen in Section 6.3.1.2, above), in that citizens should not be coerced or conditioned into providing their consent for use of their personal data for the purposes intended by the end-user under penalty of not having their signal/complaint addressed.

6.3.3.5 POWER BALANCE

Asymmetric power balances shall be avoided or mitigated when possible. This ties into the previous point – where consent is relied on, it must be made clear to data subjects that they will not suffer any negative consequences should they refuse to provide their consent, or later choose to withdraw it (in particular, it must be made clear that their signal/complaint will be duly processed regardless of this consent). Even where consent is not relied on, the end-user must ensure that it complies with all applicable legal obligations when handling citizens' personal data and must develop policies based on those data with a reasoned and critical approach, having citizens' fundamental rights and freedoms at the forefront of the decision-making process, to avoid abuse of power or arbitrariness.

6.3.4 Transparency

To ensure its compliance with the principle of transparency, as seen in 1.2.4.4 above, the end-user must ensure that it provides complete and understandable information to data subjects on their data processing practices.

Ideally, this would involve the development of an information notice, to be provided directly to citizens upon collection of their personal data, in writing. The end-user must develop such a notice with the inherent tension between the GDPR requirements of providing comprehensive information, and ensuring that the information provided is concise, transparent, intelligible and easily accessible in mind – this requires an assessment as to which information should be prioritised, what the appropriate level of detail is and which are the best means by which to convey this information to data subjects. [39] Whenever feasible, the end-user should rely on the so-called 'layered approach', allowing them to structure the information into relevant categories which the data subject can select, to ensure immediate access to the information deemed most relevant by the data subject and prevent information fatigue. [39] [68] Where information is collected outside of an online context, one way to follow this approach would be to provide citizens with an abbreviated paper-based notice at the municipality's contact centre, including a link to the more complete privacy statement made available online. [39]

Any material or substantive changes to information notices, reflecting changes to the underlying processing activities, should be communicated directly to citizens in a manner which ensures that they will be noticed. [39] It will not be valid to merely inform data subjects that they should regularly contact the municipality or check an online information notice for changes or updates, given the inherent unfairness to data subjects which this represents. [39]

6.3.5 Purpose Limitation

As seen in Section 2.2.4.5 above, to ensure compliance with the principle of purpose limitation, under Art. 5, par. 1, let. b) GDPR, the end-user must identify specific, explicit, and legitimate purposes for which personal data are to be collected and processed, and then refrain from using personal data for any other incompatible purpose. Through the presumption established for "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" in Art. 5, par. 1, let. b) GDPR, provided that the safeguards of Art. 89, par. 1 GDPR are respected – as addressed above, in Section 2.2.4.2 – it is arguably possible for the end-user to make further use of personal data collected in a commercial or healthcare context (and arguably also in the context of the municipality contact centre) to be further used for scientific research purposes (and arguably also for statistical analysis aimed at the pursuit of a public interest), provided that the appropriate safeguards of Art. 89 GDPR are in place. [40]

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.5 above can be provided on specific key elements to be considered by the end-user, under the principle of purpose limitation:

6.3.5.1 Specificity

The purposes must be specific to the processing; it should be explicitly clear to citizens why personal data is being processed.

6.3.5.2 Necessity

The end-user should, in particular, carefully assess whether the purposes for which it intends to use the platform can be used with only anonymous or aggregated data collected from signals/complaints (preferred approach), as opposed to identifiable data (e.g., specific signals/complaints linked to a specific citizen through a name or national identification number, or through other identifiers).

6.3.5.3 Compatibility

The presumption of compatibility mentioned above applies here, for research and statistical purposes.

6.3.6 Data minimisation

As seen in Section 2.2.4.6 above, compliance with the principle of data minimisation requires a minimalistic approach to personal data, in the sense that (1) as little of it as possible should be processed in order to meet an intended purpose, and (2) only personal data which are adequate, relevant and strictly necessary to meet a purpose should be used. Ultimately, if a purpose can be met without using personal data (e.g., using only anonymous or aggregated data), then no personal data should be used at all.

Considering the data sources identified by the relevant end-user for this use case [75], aside from the general guidance provided in Section 2.2.4.6 above under the principle of data minimisation, the end-user should arguably assess, for example, whether there is any added value in retaining the ability to identify the specific citizen submitting a signal/complaint through the municipality's contact centre for the policy-making purposes which are pursued. If so, the end-user should then consider whether preserving this added value serves a legitimate goal, if the added value is substantial, and if the benefits of retaining this ability outweigh the potential impact on the citizens in question. If the end-user determines that the use of personal data, preserving a link to the identity of individual citizens, is necessary, then the end-user must be able to demonstrate that each data point collected is specifically relevant to the purpose pursued. Any irrelevant personal data will be deemed as excessive and should not be collected or further processed.

6.4 Specific concerns related to ethical / societal impact

Aside from the ethical and societal concerns highlighted in Section 2.1, specific main concerns for this use case include:

1. The need to ensure accuracy of data collected and of outputs generated from such data, given the potential impact which misguided policies may have on citizens and the entire city.
2. In particular, the need to ensure that data used is representative of the city's population and does not unduly discriminate against certain groups (e.g., based on gender, race, residential neighborhood, etc.).

7 Specific issues related to use case #4 (open data policies for citizens)

In this section we will analyse in greater detail the specific ethical, legal, regulatory, and societal issues related to the fourth use case. To this regard, the main concerns to be addressed are:

1. Identification and management of requirements set by applicable labour law provisions. For example, it must be assessed whether applicable labour law provides for exceptions and/or additional safeguards on data protection legal framework.
2. The assessment of legal basis relating to the processing of special categories of data according with Art. 9 GDPR (e.g.: racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, etc.).
3. Risks related to the processing of data relating to minors.
4. Citizenship participation to policymaking initiatives.

The purpose of this use case is to assist policy makers in creating effective policies that will address employment figures. The overall goal of this is use case is for policy makers to be able to use statistics from predictive algorithms from the toolkit to assist in making decision during policy creation process. The main objective will be to design the algorithms that will help predict future trends using the unemployment datasets provided.

In more detail, the objectives of the use case are:

1. To enable users to use the platform to assist with the policy creation process.
2. To identify relevant KPIs using the PolicyCLOUD platform.
3. To highlight the correlation and information to help with decision making.

In the context of the use case, the policy makers will:

1. Conduct analysis based on the statistics relating to specific time periods. To this regard, the goal is to use the analytics and visualizations produced by the PolicyCLOUD platform to identify key information that could help determine groups of citizens that are affected by unemployment.
2. Use predictive analytics to predict a future outcome.
3. Identify trends in specific age groups. Once a particular trend has been identified, then an appropriate policy can be designed based on what has been learnt from the data. [49] [69]

The related use case scenarios are:

1. Conducting analysis based off the statistics on specific time periods. For example, the unemployment is expected to go up during the year 22 due to the current pandemic. Therefore, the statistics recorded against the current year can help to identify the possible unemployment rate if there is second wave of infections the following year. The goal of this scenario is to use the analytics and visualisations produced from the PolicyCLOUD platform to identify key information that could help determine groups of citizens that are affected by unemployment.

2. Using predictive analysis to predict a future outcome. The goal of this scenario is to use specially designed algorithms from PolicyCLOUD to predict future outcomes.

3. Factors such as age, gender, and time-based statistics such as month/ year will be key indicators to highlight common trends in specific age groups. Summaries of the total amount of unemployed citizens within a specific time can be a key component to identifying significant trends. The goal for this use case scenario is to identify trends in specific age groups. Once a particular trend has been identified then an appropriate policy can be designed based on what has been learnt from the data.

The common goals of the above described scenarios are to allow policy makers to be able to use statistics from predictive algorithms from the toolkit to assist in making decision during policy creation process.

7.1 Compliance Assessment around Selection of Data Sources

As noted in Section 2.2 above, applicable legal, regulatory, ethical and/or societal requirements may restrict an end-user's leveraging of a data source which might otherwise be considered appropriate from a practical perspective (i.e., useful towards the goal which the PolicyCLOUD end-user wishes to achieve).

For this use case, the following data source has been indicated as relevant [75]:

1. Unemployment Claimant Count (Dataset 20), described as "unemployment open data", consisting of "different types of benefits, which include registered individuals jobseekers Allowance (JSA) claimants and Universal Credit claimants who are actively seeking work", collected from the "Open Data Camden" platform.

The end-user must carry out a comprehensive assessment, following the guidelines noted in Section 3.1.2.1 above, to ensure that this data source can be adequately leveraged, from a legal, ethical, and societal perspective. In particular, the following key points should be considered:

Dataset 19

1. Given that the end-user is presumably responsible for management of this database, there should in principle be no contractual restrictions towards leveraging information obtained from the Open Data Camden Platform for the purposes of the end-user. If this is not the case, then the end-user must assess whether the contractual terms applicable to the platform might prevent this – in particular, the Open Government License for public sector information provided by the Open Data Camden platform[106] – to ensure that data collected from Open Data Camden can be leveraged for the purposes of the end-user.

2. On this point, the terms of the Open Government License state that users accepting it are free to "copy, publish, distribute and transmit the Information", "adapt the Information" and "exploit the Information commercially and non-commercially for example, by combining it with other information, or by including it in your own product or application". This is subject to a need to "acknowledge the source of the Information in your product or application by including or linking to any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence", as well as the need to include a fixed attribution in case an "Information Provider" does not do so. However, exceptions exist, including "personal data in the Information". This broad license suggests that the end-user is generally authorised to use Open Camden Data as intended by the use case, albeit with the possible exclusion of any personal data included therein.

3. While this is not clear, it is possible that leveraging this data source may imply the processing of personal data. The obligations around use of this data source, in case any personal data are processed, are further developed in Section 7.2 below.

4. The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete (also in terms of representativeness of the population of the borough), considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

7.2 Specific concerns on privacy and data protection

In this Section, we will present an overview of the main issues which the end-user – acting as a controller – must bear in mind in the definition of requirements for use of the platform in each of the presented scenarios, with reference to the GDPR's data protection principles described in Section 2.2.4 above. Only those principles which present specific additional concerns to those provided in the general sections will be addressed – if a specific principle is not covered for this use case, the general section on that principle applies.

It is not, yet, clear whether the data source indicated for this use case [75] presents a potential for capturing personal data. As such, the following sub-sections should be borne in mind only to the extent that any personal data (i.e., information about identified, or identifiable individuals) is to be extracted and further processed from this data source.

[106] Available at: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

7.2.1 Lawfulness

As noted in Section 2.2.4.1 above, any use of personal data must be performed on the basis of consent provided by the individuals whose data is used ("data subjects"), or otherwise on some other legitimate basis laid down in law, as set out in the GDPR or in other Union or Member State laws referred to by the GDPR.

To this end, considering the data source identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.1 above can be provided on the particularities of specific legal bases which may potentially be relevant for consideration:

7.2.1.1 LEGAL OBLIGATION

As noted in Section 2.2.4.1 above, one legal basis which may potentially be applicable is the need to process personal data to comply with a legal obligation to which the end-user is subject[107]. It is recommended, as a first step in the lawfulness assessment, for the end-user to assess whether this legal basis may apply to the processing activities it may envision performing through PolicyCLOUD – i.e., if the requirements described in Section 2.2.4.1 above for this legal basis are met for this specific use case – before considering possible alternatives.

7.2.1.2 CONSENT

As noted in Section 2.2.4.1 above, where the above legal basis is unavailable, another which may apply is consent provided by data subjects for use of their personal data, for the purpose of generating aggregated information from which the end-user may be able to draw insights into relevant trends and issues arising at the borough-level, so as to use those insights to improve the decision-making process around policies on unemployment. This legal basis may be particularly relevant for the data source in question, to the extent that it would be feasible to ask individuals applying for the relevant types of benefits to consent to further processing of their personal information for such purposes.

In order to obtain valid consent from citizens for this further processing of their personal data, the end-user should consider the feasibility of requesting consent from citizens when they decide to submit an application for such benefits, in a manner which meets all of the consent requirements of the GDPR (as described in Section 2.2.4.1 above). In particular, the end-user should bear in mind that making this consent mandatory in order to allow data subjects to benefit from the provision of a service (e.g., where a citizen would not be allowed to apply for a specific benefit without consenting to further processing of their personal data for analytics purposes) would affect the freedom of their consent and, therefore, its validity. Given the inherent imbalance of power between the end-user for this use case (a public authority) and an individual citizen, it must be made absolutely clear to citizens that they will suffer no adverse consequences if they choose not to provide their consent, or to later withdraw it – in particular, they must be clearly informed that their application for benefits will still be duly processed, even if they do not provide this additional consent.

[107] Art. 6, par. 1, let. c) GDPR.

7.2.1.4 PUBLIC INTEREST

As noted in Section 2.2.4.1 above, where neither of the above two options are available, the end-user may consider whether it can justify the intended processing of personal data on the need to perform a task carried out in the public interest, or in the exercise of official authority[108].

Given that it is not possible for a public authority to rely on the legal basis set out in Art. 6, par. 1, let. f) GDPR in the performance of their tasks, this legal basis presents the most flexible approach available to public authorities under those circumstances, and thus the end-user is recommended to assess whether, for this use case, the requirements for this specific legal basis – as described in in Section 2.2.4.1 above – are met.

7.2.2 Lawfulness (Special Categories of Personal Data)

As noted in Section 2.2.4.2 above, where any special categories of personal data are to be collected and further processed, an applicable derogation to the GDPR's general prohibition on the processing of these personal data[109], from those listed in Art. 9, par. 2 GDPR, or as may be further provided under applicable Member State law[110], must also be identified. For clarity, to lawfully process special categories of personal data, a controller must identify an applicable legal basis under Art. 6 GDPR AND an applicable derogation under Art. 9 GDPR.

To this end, considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.2 above can be provided on the particularities of specific derogations which may potentially be relevant for consideration:

7.2.2.1 EXPLICIT CONSENT

On this point, we refer to the requirements set out in Section 2.2.4.1 above, as all of those must be met to ensure validity of consent, and on the general observations as to the viability of reliance on consent for this use case made in Section 7.2.1.2 above.

7.2.2.2 SUBSTANTIAL PUBLIC INTEREST

As noted in Section 2.2.4.2 above, this derogation requires a basis in EU or Member State law applicable to the controller which must:

1. Be proportionate to the interest pursued.
2. Respect the essence of the right to data protection.
3. Provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests.

The end-user should assess whether this derogation may be applicable, considering the purposes which it may be seeking to pursue through processing of personal data via the platform.

[108] Art. 6, par. 1, let. e) GDPR.
[109] Established by Art. 9, par. 1 GDPR.
[110] In particular, regarding the processing of genetic data, biometric data or data concerning health, as set out in Art. 9, par. 4 GDPR.

7.2.2.4 STATISTICAL PURPOSES

In light of the requirements explained in Section 2.2.4.2 above, as the processing of special categories of personal data, if relevant, may be carried out in this use case for statistical purposes, the end-user should determine whether this processing can be based on EU or Member State law applicable to the end-user, which must:

1. Be proportionate to the interest pursued.
2. Respect the essence of the right to data protection.
3. Provide for suitable and specific measures to safeguard data subjects' fundamental rights and interests, this derogation may be applicable.

In this case, the end-user and PolicyCLOUD must collaborate to ensure that the specific safeguards further specified in Art. 89, par. 1 GDPR, as described in Section 2.2.4.2 above, can be implemented.

7.2.3 Fairness

As noted in Section 2.2.4.3 above, for the processing to be considered as fair under the GDPR, the end-user shall ensure that personal data are handled in ways that may be reasonably expected by data subjects and not use such data in a way that may produce unjustified adverse effects on them.

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.3 above can be provided on specific key elements to be considered by the end-user, under the principle of fairness:

7.2.3.1 EXPECTATION

If data subjects are led to believe that personal data collected on them will be used to improve the borough's policy-making abilities, this should be the only objective pursued with those personal data – using them to profile and target individuals, or for other unrelated and arguably illegitimate purposes (e.g., sending of marketing communications), must be strictly avoided. This will imply controls around purpose limitation, including access control.

7.2.3.2 NON-DISCRIMINATION

The controller shall not discriminate against data subjects. In particular, personal data should not be collected on citizens for the purpose of discriminating against them (such as to cause harm or detriment to citizens belonging to certain disadvantaged or minority groups), nor should this be the end-result of policies developed using citizens' personal data – this requirement is strongly tied to applicable ethical considerations of avoidance of bias and non-discrimination, as seen in Section 2.1.3 and 2.1.4 above.

7.2.3.3 NON-EXPLOITATION

The controller shall not exploit the needs or vulnerabilities of data subjects. Considering the inherent imbalance of power between the controller (a public authority) and individual citizens, this is particularly relevant when assessing the freedom of consent, where this legal basis is leveraged (as seen in Section 7.2.1.2, above), in that citizens should not be coerced or conditioned into providing their consent for use of their personal data for the purposes intended by the end-user under penalty of not having their application for benefits duly processed.

7.2.3.5 POWER BALANCE

Asymmetric power balances shall be avoided or mitigated when possible. This ties into the previous point – where consent is relied on, it must be made clear to data subjects that they will not suffer any negative consequences should they refuse to provide their consent, or later choose to withdraw it (in particular, it must be made clear that their application for benefits will be duly processed regardless of this consent). Even where consent is not relied on, the end-user must ensure that it complies with all applicable legal obligations when handling citizens' personal data and must develop policies based on those data with a reasoned and critical approach, having citizens' fundamental rights and freedoms at the forefront of the decision-making process, to avoid abuse of power or arbitrariness.

7.2.4 Transparency

To ensure its compliance with the principle of transparency, as seen in Section 2.2.4.4 above, the end-user must ensure that it provides complete and understandable information to data subjects on their data processing practices.

Ideally, this would involve the development of an information notice, to be provided directly to citizens upon collection of their personal data, in writing. The end-user must develop such a notice with the inherent tension between the GDPR requirements of providing comprehensive information, and ensuring that the information provided is concise, transparent, intelligible and easily accessible in mind – this requires an assessment as to which information should be prioritised, what the appropriate level of detail is and which are the best means by which to convey this information to data subjects. [39] Whenever feasible, the end-user should rely on the so-called 'layered approach', allowing them to structure the information into relevant categories which the data subject can select, to ensure immediate access to the information deemed most relevant by the data subject and prevent information fatigue. [39] [68] Where information is collected outside of an online context, one way to follow this approach would be to provide citizens with an abbreviated paper-based notice upon submission of an application for benefits, including a link to the more complete privacy statement made available online. [39]

Any material or substantive changes to information notices, reflecting changes to the underlying processing activities, should be communicated directly to citizens in a manner which ensures that they will be noticed. [39] It will not be valid to merely inform data subjects that they should regularly contact the end-user or check an online information notice for changes or updates, given the inherent unfairness to data subjects which this represents. [39]

7.2.5 Purpose Limitation

As seen in Section 2.2.4.5 above, to ensure compliance with the principle of purpose limitation, under Art. 5, par. 1, let. b) GDPR, the end-user must identify specific, explicit, and legitimate purposes for which personal data are to be collected and processed, and then refrain from using personal data for any other incompatible purpose. Through the presumption established for "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes" in Art. 5, par. 1, let. b) GDPR, provided that the safeguards of Art. 89, par. 1 GDPR are respected – as addressed above, in Section 2.2.4.2 – it is arguably possible for the end-user to make further use of personal data collected in a commercial or healthcare context (and arguably also in the context of the filing of benefits applications in Camden) to be further used for scientific research purposes (and arguably also for statistical analysis aimed at the pursuit of a public interest), provided that the appropriate safeguards of Art. 89 GDPR are in place. [40]

Considering the data sources identified by the relevant end-user for this use case [75], further guidance to complement that provided in Section 2.2.4.5 above can be provided on specific key elements to be considered by the end-user, under the principle of purpose limitation:

7.2.5.1 Specificity

The purposes must be specific to the processing; it should be explicitly clear to citizens why personal data is being processed.

7.2.5.2 Necessity

The end-user should, in particular, carefully assess whether the purposes for which it intends to use the platform can be used with only anonymous or aggregated data collected from benefits applications (preferred approach), as opposed to identifiable data (e.g., specific applications linked to a specific citizen through a name or national identification number, or through other identifiers).

7.2.5.3 Compatibility

The presumption of compatibility mentioned above applies here, for research and statistical purposes.

7.2.6 Data minimisation

As seen in Section 2.2.4.6 above, compliance with the principle of data minimisation requires a minimalistic approach to personal data, in the sense that (1) as little of it as possible should be processed in order to meet an intended purpose, and (2) only personal data which are adequate, relevant and strictly necessary to meet a purpose should be used. Ultimately, if a purpose can be met without using personal data (e.g., using only anonymous or aggregated data), then no personal data should be used at all.

Considering the data source identified by the relevant end-user for this use case [75], aside from the general guidance provided in Section 2.2.4.6 above under the principle of data minimisation, the end-user should arguably assess, for example, whether there is any added value in retaining the ability to identify the specific citizen submitting an application for benefits for the policy-making purposes which are pursued. If so, the end-user should then consider whether preserving this added value serves a legitimate goal, if the added value is substantial, and if the benefits of retaining this ability outweigh the potential impact on the citizens in question. If the end-user determines that the use of personal data, preserving a link to the identity of individual citizens, is necessary, then the end-user must be able to demonstrate that each data point collected is specifically relevant to the purpose pursued. Any irrelevant personal data will be deemed as excessive and should not be collected or further processed.

7.3 Specific concerns related to ethical / societal impact

Aside from the ethical and societal concerns highlighted in Section 2.1, specific main concerns for this use case include:

1. The need to ensure accuracy of data collected and of outputs generated from such data, given the potential impact which misguided policies may have on citizens and the entire borough.

2. In particular, the need to ensure that data used is representative of the borough's population and does not unduly discriminate against certain groups (e.g., based on gender, race, residential neighborhood, etc.).

8 Implementation of the ethical, legal, societal, and regulatory requirements in the solutions

8.1 Application of the compliance by design principle

In this section we will analyse how, during the development of PolicyCLOUD, the compliance with the ethical, legal, societal, and regulatory requirements identified in the above sections will be assessed.

With specific regards to data protection and privacy issues, we will also define a methodology for the implementation of a DPIA which will be conducted with regards to each of the four pilots.

8.1.1 Compliance by design approach

To address all the relevant ethical, legal, regulatory, and societal risks related to the project, a compliance by design approach shall be adopted.

Compliance by design means applying a systematic approach to integrating regulatory requirements into tasks and processes. The effective implementation of this principle will be based on the detailed and structured analysis of all the applicable requirements (as identified in the previous sections), followed by translation of rules into compliance processes. [74]

To do so, a three-stage approach may will be applied:

1. The first stage will be dedicated to the identification and the assessment of regulatory requirements.
2. The second stage will include the analysis on how the rules apply to individual processes.
3. The first stage will focus on the design and implementation of a roadmap.

8.1.2 Data protection and privacy

To address the issues related to the Project and the use cases concerning data protection and privacy, DPIAs will be implemented and the results of the same will be presented in the context of deliverables D3.6 and D3.9 of the Project. Through the DPIAs, PolicyCLOUD will assess the processing operations to be performed, as well as the technologies, tools, and systems to be used, to identify inherent risks in a structured manner. Furthermore, the DPIAs will be used to identify measures which can be implemented to bring those risks down to acceptable levels. The DPIA reports will contain a systematic description of the envisaged processing operations, the purposes for which personal data will be processed, an assessment of the legitimate interests pursued (where applicable), an assessment of the necessity and proportionality of the operations in relation to those purposes, an assessment of the risks to the rights and freedoms of data subjects, and a description of the measures envisaged to address those risks, as noted in Art. 35, par. 7 GDPR. [75]

The DPIAs will be performed according to the methodology defined in the international standard ISO/IEC 29134.

The process for the performance of the DPIAs will include:

1. A preparation phase, during which the DPIA teams will be set and provided with direction, the DPIA plan will be prepared, the necessary resources will be determined, and the relevant stakeholders will be engaged.

2. A performance phase, during which, the information flows of personal data will be identified, the implication of the Project or the use case will be analysed, the relevant privacy safeguarding requirements will be determined, the data protection and privacy risk will be assessed, a risk treatment plan will be defined.

3. A follow up phase, during which a DPIA report will be prepared and published and the risk treatment plan will be implemented. In this context, also a review and/or reaudit program of the DPIA will be defined, to monitor both the correct implementation of the risk treatment plan and of the potential changes to the previously assessed personal data processing activities.

In the context of the DPIAs, adequate attention will be also dedicated to the topic of the security of personal data. To this regard, the relevant threats to availability, confidentiality and integrity of personal data will be duly analysed, including:

1. Threats related to the supporting assets on which the personal data relies, such as:
    - User provided hardware and software (e.g., smartphones, tablets, internet browser software, etc.).
    - Generic hardware (e.g., computers, communications relay, USB drives, hard drives, etc.).
    - Software (operating systems, messaging, databases, business applications, etc.)
    - Computer channels (e.g., cable, wireless, fibre optic, etc.).
    - Individuals (e.g., users, administrators, top management, etc.).
    - Paper documents (e.g., printing, photocopying, etc.).
    - Paper transmission channels (e.g., mail, workflow, etc.).
2. Threats related to the actions of those supporting assets, such as:
    - Abnormal use and/or function creep, in which supporting assets are diverted from their intended context of use without being altered and/or damaged.
    - Damaging, in which supporting assets are completely or partially damaged.
    - Espionage, in which supporting assets are observed without being damaged.
    - Loss, in which supporting assets are lost, stolen, sold, or given away, so it is no longer possible to exercise property rights.
    - Modification and/or change, in which supporting assets are transformed.
    - Overload and/or exceeded limits of operations, in which supporting assets are overloaded, overexploited and/or used under conditions not permitting them to function properly. [75]

8.2 Compliance checklist

In this section, on the basis of the analysis previously performed in the context of this deliverable, we summarize in a checklist presented in TABLE 1 below a list of key compliance controls which shall be respected in order to ensure the ethical, legal, regulatory and societal sustainability of the Project.

Area	Type of requirement	Control
Ethical/Societal	Ethical and societal requirements related to cloud computing	It is necessary to have a clear and specific framework in place with the IaaS provider, in which objectives, processes and results expected from PolicyCLOUD platform are clearly set out, so as to specify the capabilities needed from the IaaS provider.
Ethical/Societal	Ethical and societal requirements related to cloud computing	The framework developed with the IaaS provider must define appropriate service levels so as to ensure that the platform and its data will be kept promptly available to PolicyCLOUD and end-users, identifying a maximum amount of acceptable service downtime and ensuring the possibility to recover data which may be lost during the interruption. Infringement of these levels should preferably be subjected to appropriate contractual penalties.
Ethical/Societal	Ethical and societal requirements related to cloud computing	The environmental impact of the infrastructure necessary to support the functioning of the cloud-based system is reduced to a minimum.
Ethical/Societal	Ethical and societal requirements related to big data management	The platform and its end-users must weigh the interests of the data subjects appropriately and find effective means to provide information about the activities performed on personal data, considering also the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the use case.
Ethical/Societal	Ethical and societal requirements related to big data management	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely test the analytics components of the platform to ensure that they do not skew knowledge obtained from data in a biased manner.
Ethical/Societal	Ethical and societal requirements related to big data management	The platform should incorporate adequate technical and organisational security measures, developed as a result of a dedicated security risk assessment targeting potential threats generated in particular from reliance on big data analysis.
Ethical/Societal	Ethical and societal requirements related to AI	Mechanisms facilitating the auditability of AI systems (e.g., traceability of the development process, the sourcing of training data, and the logging of the AI system processes, outcomes, and positive and negative impacts) should be put in place.
Ethical/Societal	Ethical and societal requirements related to AI	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.
Legal/Regulatory	Contractual protection of data sources	Should a selected data source be subject to contractual terms which prevent its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the data source owner.

Area	Type of requirement	Control
Legal/Regulatory	Legal protection of databases	Should a selected data source be eligible for database copyright or sui generis right protection which prevents its use as intended by PolicyCLOUD, that data source should not be registered on PolicyCLOUD without proper authorisation from the rights holder.
Legal/Regulatory	Protection of copyright	Should a selected data source, or a relevant part of that data source, which is to be extracted, be eligible for copyright which prevents its use as intended by the PolicyCLOUD user, that data source should not be registered on PolicyCLOUD without proper authorisation from the rights holder.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD must assess which of the legal bases afforded by the GDPR may be applicable and implementable for an intended processing of personal data. This assessment must consider the full context of the processing activities which are intended, including the specific data sources to be used and the specific goals to be reached using the platform.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD should only handle personal data in ways that may be reasonably expected and not use such data in a way that may produce unjustified adverse effects on data subjects.
Legal/Regulatory	Personal data protection and privacy	When Data Subjects seek to exercise their rights granted by the GDPR or other applicable data protection laws PolicyCLOUD shall be capable to facilitate the exercise of these rights.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD and the end-users must be clear and honest with Data Subjects about the identity of the data controller which is collecting, processing and storing personal data, the methods used to process personal data, and the purposes of processing.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD and the end-users, in relation to the processing activities which they may respectively perform, as controllers, shall lay down a specific and easily accessible document which duly informs Data Subjects of the processing activities carried out in the context of the Project: a privacy policy.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD shall made available the Privacy Policy on its cloud-based platform, with appropriate steps taken to make it available to the Data Subjects whose personal data are used in the context of the Project. End-users should likewise ensure that the above information is available to Data Subjects on public websites under their control.
Legal/Regulatory	Personal data protection and privacy	Platform users shall be limited both from a technical and from a contractual point of view in how they can process personal data which are collected and managed through the PolicyCLOUD platform.
Legal/Regulatory	Personal data protection and privacy	Internal policies shall be implemented to make users aware of what they can and cannot do with the personal data collected for the different use cases and more in general for the execution of the Project.

Area	Type of requirement	Control
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD should implement technical measures, including hashing and cryptography, to limit the possibility of repurposing personal data.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD shall only collect personal data, which is adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. As such, for each purpose of processing connected to the Project, it shall identify the minimum amount of personal data needed to fulfil such purpose.
Legal/Regulatory	Personal data protection and privacy	End-users and PolicyCLOUD must ensure that appropriate steps are taken to verify the accuracy of any personal data collected, to maintain those personal data up to date over time, and to allow data subjects to correct, complete or update their own personal data when needed.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD shall implement technical and organisational measures aimed at guaranteeing the accuracy and quality of personal data included in the cloud-based platform and shall provide means to Data Subjects for contributing to the maintenance of data that is always accurate and up-to-date.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD shall keep personal data in a form which permits identification of Data Subjects for no longer than is necessary for the purposes for which the personal data are processed. Even where personal data are collected in a fair and lawful manner, they cannot be stored for longer than actually needed, unless a reason for further processing exists, and provided that a legal basis for such further processing has been detected by PolicyCLOUD pursuant to the purpose limitation principle. Therefore, PolicyCLOUD shall proceed to the erasure of personal data from the cloud-based platform when it has no reasons for keeping them or, alternatively it shall anonymize and aggregate such data.
Legal/Regulatory	Personal data protection and privacy	It should be demonstrable that the security measures implemented on the platform were chosen as a result of a documented risk assessment, with justifications as to why those measures were deemed adequate to address the specific risks identified.
Legal/Regulatory	Personal data protection and privacy	There should be clearly defined rules and specific channels on the reporting of security incidents or abnormal events related to the platform., and all persons working with the platform should be made aware of the types of occurrences which may qualify as a reportable security incident.
Legal/Regulatory	Personal data protection and privacy	The processes implemented to address personal data breaches by PolicyCLOUD should ensure that all relevant information on a personal data breach and the manner in which it was handled is documented in a register of personal data breaches, as set out in Art. 33, par. 5 GDPR, including all facts pertaining to the personal data breach, its effects and remedial action taken, including notifications to end-users, supervisory authorities and/or Data Subjects, as well as all technical

Area	Type of requirement	Control
		and organizational mitigation measures applied, documented assessments carried out, including those performed to classify the incident as a personal data breach, as well as to classify a personal data breach in terms of category and severity level. Post-breach analyses should also be carried out, to validate the effectiveness of the breach management process, identify areas of improvement, and identify, based on a root cause analysis of the incident, adequate technical and organisational measures to reduce or eliminate the likelihood of recurrence.
Legal/Regulatory	Personal data protection and privacy	A security risk assessment, as part of an overall DPIA, should be carried out, to identify possible threats and risks to the fundamental rights, freedoms and interests of Data Subjects and the specific security measures implemented or which should be implemented to address them.
Legal/Regulatory	Personal data protection and privacy	PolicyCLOUD and the end-users shall take responsibility for what is done with personal data and how it complies with the personal data protection principles, implementing measures, documents and records to demonstrate that appropriate processes and procedures are in place to ensure that personal data are collected, processed and stored in such a way that is compliant with the GDPR and with other applicable data protection laws.
Legal/Regulatory	Requirements related to the use of the Cloud Infrastructure	A DPA shall be executed between PolicyCLOUD and the IaaS provider, including all the requirements defined by Art. 28 GDPR.
Ethical/Legal/Regulatory/Societal	Requirements related to the use of the Cloud Infrastructure	PolicyCLOUD shall perform a comprehensive risk assessment, focused on the likelihood and impact of threats to confidentiality, integrity and/or availability of assets stored on the platform, and to the resilience of the infrastructure of the platform and systems itself, as well as relevant compliance obstacles raised by use of the cloud infrastructure, and assess whether the technical and organisational measures put in place by the IaaS provider in relation to the cloud infrastructure sufficiently mitigate any relevant risks identified.
Ethical/Legal/Regulatory/Societal	Requirements related to the use of the Enhanced Interoperability component	Regarding the Enhanced Interoperability component, PolicyCLOUD should ensure that enough testing is performed to ensure a reasonable degree of statistical accuracy for annotations and connections established. This may require an extensive training exercise involving various kinds of data sources, to refine rules used by this component for these activities.
Ethical/Legal/Regulatory/Societal	Requirements related to the use of the Enhanced Interoperability component	The Enhanced Interoperability component shall methodically log the activities performed to ensure traceability and presents correlations established and their rationale to the end-user for confirmation, to provide some level of human validation of those correlations. End-users should also be advised of the possibility of false positive correlations, so that they are incentivised to verify the validity of correlations made.

Area	Type of requirement	Control
Ethical/Legal/Regulatory/Societal	Requirements related to data analytics	PolicyCLOUD should ensure that enough testing is performed to ensure a reasonable degree of statistical accuracy for the analytics functions. Any operations performed on data should be methodically logged to ensure traceability, and the general rationale and logic behind the analytics performed should be explained to end-users, so this can be considered during their decision-making process. End-users should be advised of the possibility of false positives or false negatives and errors in result presentation, so that they are incentivised to critically examine results produced by the analytics functions in their decision-making process.
Ethical/Legal/Regulatory/Societal	Requirements related to the PDT	It should be defined terms and conditions for the use of the PDT, to properly regulate the service relationship established between PolicyCLOUD and the end-user or the organisation to which the end-user belongs. These terms and conditions would need to be accepted for the use of the PDT to be allowed.
Ethical/Legal/Regulatory/Societal	Requirements related to the PDT	End-users shall be enabled to understand the output presented to them by the PDT.
Ethical/Legal/Regulatory/Societal	Requirements related to use case #1	Children inclusion into the investigation tool activities must be avoided, excluding collection of data related to minors, if possible, or storage after accidental collection.
Ethical/Legal/Regulatory/Societal	Requirements related to use case #1	The Consortium shall carry out an assessment of the impact of imagined processing operations on the protection of personal data and ethical values, to identify and reduce the privacy risks and the likeliness of rights and freedoms infringements.
Ethical/Legal/Regulatory/Societal	Requirements related to use case #2	The end-user should identify steps to reasonably assure itself of the reliability of the data sources, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete, also in terms of representativeness of the population of the borough, considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.
Ethical/Legal/Regulatory/Societal	Requirements related to use case #3	The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete, considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.
Ethical/Legal/Regulatory/Societal	Requirements related to use case #4	The end-user should identify steps to reasonably assure itself of the reliability of this data source, in terms of the likelihood that any data collected from the data source may be false, inadequate, inaccurate or incomplete, considering the purpose for which the data source is to be used. The end-user should identify, and document specific steps taken to address reliability and accuracy concerns detected.

TABLE 1 – COMPLIANCE CHECKLIST

9 Conclusion and Next Steps

In this deliverable, we have set out to present the general legal/regulatory and ethical/societal concerns which the development of the PolicyCLOUD platform may face, in Section 2. These general concerns were applied to specific situations, namely the platform's foreseen components (to the extent that information was available, as of the date of this deliverable, on each component), in Section 3, and each of the four use cases currently envisioned, in Sections 4 to 7. This deliverable further sought to identify, in Section 8, practical requirements, based on those concerns, which may be implemented into the design of the platform (and should be considered by end-users regarding their intended use of the platform).

Our focus concerning legal/regulatory concerns and requirements, in this iteration of this framework, was on requirements borne out of Union law, as opposed to local laws applicable in different jurisdictions/Member States. Furthermore, as of the date of this deliverable, several, if not most components of the PolicyCLOUD platform are under development (not least of which is the Data Marketplace). The four uses cases also continue to be refined, with end-users defining further relevant data sources and use case scenarios. As such, all concerns and requirements reflected in this deliverable were provided based on the current state of the Project's definition, as of the date of completion of this deliverable.

In subsequent iterations (D3.6 and D3.9, due on M22 and M34 respectively), to the extent that the development of the platform suggests that this may be relevant and necessary, this deliverable will be expanded to include an analysis of additional or particular concerns relevant at a local level – considering, in particular, the jurisdictions in which each of the end-users of the currently envisioned four use cases are established. Additionally, subsequent iterations of this deliverable will be duly updated to reflect all these relevant developments, with sections being expanded on or amended as needed. Finally, in an effort to increase the effectiveness of this framework, subsequent iterations will rely on feedback provided by the Consortium to improve on the practical requirements identified (including in terms of feasibility of their implementation), further tailoring them to the specificities of the platform and use cases. It is expected that the final iteration of this framework (D3.9) will reflect a set of requirements which are understood and accepted by the Consortium as vital to ensure the ethical and legal soundness of the Project.

References

[1] EUROPEAN DATA PROTECTION SUPERVISOR, *Opinion 4/2015, Towards a new digital ethics: Data, Dignity and Technology*, https://edps.europa.eu/data-protection/our-work/publications/opinions/towards-new-digital-ethics-data-dignity-and_en, retrieved 2020-12-19.

[2] Milchram C., van de Kaa G., Doorn N. & Künneke R. (2018), *Review Moral Values as Factors for Social Acceptance of Smart Grid Technologies*, https://www.mdpi.com/2071-1050/10/8/270310 (retrieved 2020-12-06).

[3] Taebi B. (2017), *Bridging the Gap between Social Acceptance and Ethical Acceptability*, Risk Analysis, 37(10), 1817-1827.

[4] EUROPEAN COMMISSION, *A European Strategy for Data,* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy, retrieved 2020-12-18.

[5] Philippon T. (2019), *The Great Reversal. How America Gave Up on Free Markets (1)*, Belknap Press, Cambridge.

[6] Foutty J. & Bawa R. (2018), *The Societal Impact of Cloud*, https://deloitte.wsj.com/cio/2018/11/14/the-societal-impact-of-cloud/, retrieved 2020-11-28.

[7] SWIPO, *Swipo – The association on switching and porting*, https://swipo.eu/, retrieved 2020-12-18.

[8] CSPCERT EUROPE, *European trusted cloud service provider working group*, See https://cspcerteurope.blogspot.com/, retrieved 2020-12-18.

[9] ENISA, Cloud security, https://www.enisa.europa.eu/topics/cloud-and-big-data/cloud-security, retrieved 2020-12-19.

[10] EUROPEAN COMMISSION, *Practical guidance for businesses on how to process mixed datasets*, https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets, retrieved 2020-12-19.

[11] Himma K.E. (2007), *Foundational Issues in Information Ethics*, Library Hi Tech, 25(1), 79–94.

[12] Owen R., Bessant J.R. & Heintz M. (eds.) (2013), *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, John Wiley & Sons, West Sussex.

[13] Sheppard A.F. (2016), *Developing Model Cloud Computing Contracts—Research Proposal*, https://interparestrust.org, retrieved 2020-12-06.

[14] EUROPEAN COMMISSION, *Study on the Economic Detriment to Small and Medium-Sized Enterprises Arising from Unfair and Unbalanced Cloud Computing Contracts*, https://ec.europa.eu/info/publications/study-economic-detriment-small-and-medium-sized-enterprises-arising-unfair-and-unbalanced-cloud-computing-contracts_en, retrieved 2020-12-19.

[15] Lytras M.D. & Visvizi A. (2019), *Big Data and Their Social Impact: Preliminary Study*, https://www.mdpi.com/20171-1050/11/18/5067, retrieved 2020-11-28.

[16] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 3/2013 on purpose limitation*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf, retrieved 2020-12-19.

[17] IBM, *Explore IBM software and solutions,* https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html, retrieved 2020-12-19.

[18] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 05/2014 on Anonymisation Techniques*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, retrieved 2020-12-19.

[19] EUROPEAN DATA PROTECTION SUPERVISOR, *Opinion 8/2016 on coherent enforcement of fundamental rights in the age of big data*, https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf, retrieved 2020-12-19.

[20] EUROPEAN DATA PROTECTION SUPERVISOR, *Opinion 7/2015, Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability*, https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf, retrieved 2020-12-19.

[21] Schneier B. (2016), *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World (1)*, W. W. Norton & Company, New York.

[22] OSLO DATA PROTECTION AUTHORITY, *Big data – privacy principles under pressure,* https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf, retrieved 2020-12-19.

[23] EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, *The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context*, https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data, retrieved 2020-12-19.

[24] Harris L. & Harrigan P. (2015), *Social Media in Politics: The Ultimate Voter Engagement Tool or Simply an Echo Chamber?*, Journal of Political Marketing, 14(3), 251-283.

[25] INFORMATION COMMISSIONER'S OFFICE, *Investigation into the use of data analytics in political campaigns. Investigation update*, https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf, 2020-12-19.

[26] HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Ethics Guidelines for Trustworthy AI*, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, retrieved 2020-12-19.

[27] EUROPEAN COMMISSION, *The EU Cybersecurity Act,* https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-act, retrieved 2020-12-19.

[28] HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment, retrieved 2020-12-19.

[29] EUROPEAN COMMISSION, *Pilot the Assessment List of the Ethics Guidelines for Trustworthy AI*, https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0, retrieved 2020-12-19.

[30] EUROPEAN GROUP ON ETHICS IN SCIENCE AND NEW TECHNOLOGIES, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf, retrieved 2020-12-19.

[31] Floridi L., Cowls J., Beltrametti M., Chatila R., Chazerand P., Dignum V., Luetge C., Madelin R., Pagallo U., Rossi F., Schafer B., Valcke P. & Vayena E.J.M. (2018), *AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Minds and Machines*, 28(4), 689-707.

[32] EUROPEAN DATA PROTECTION SUPERVISOR, *Public Consultation on Digital Ethics. Summary of outcomes*, https://edps.europa.eu/data-protection/our-work/publications/reports/public-consultation-digital-ethics-summary-outomes_en, retrieved 2020-12-19.

[33] EUROPEAN IPR HELPDESK, Fact sheet on *Copyright Essentials*, https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Fact-Sheet-copyright_essentials.pdf, retrieved 2020-12-19.

[34] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm, retrieved 2020-12-19.

[35] EUROPEAN DATA PROTECTION BOARD, *Guidelines 05/2020 on consent under Regulation 2016/679,* https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf, retrieved 2020-12-19.

[36] EUROPEAN DATA PROTECTION BOARD, *Guidelines 8/2020 on the targeting of social media users*, https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/guidelines-082020-targeting-social-media-users_en, retrieved 2020-12-19.

[37] EUROPEAN DATA PROTECTION BOARD, *Guidelines 06/2020 on the interplay of the Second Payment Services Directive and the GDPR*, https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/guidelines-062020-interplay-second-payment-services_en, retrieved 2020-12-19.

[38] EUROPEAN DATA PROTECTION BOARD, *Guidelines 4/2019 on Article 25 – Data Protection by Design and by Default*, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en, retrieved 2020-12-19.

[39] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Transparency under Regulation 2016/679*, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227, retrieved 2020-12-09.

[40] EUROPEAN DATA PROTECTION SUPERVISOR, *Preliminary Opinion on data protection and scientific research*, https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en, retrieved 2020-12-10.

[41] EUROPEAN UNION AGENCY FOR CYBERSECURITY, *Technical Guidelines for the implementation of minimum security measures for Digital Service providers*, https://www.enisa.europa.eu/publications/minimum-security-measures-for-digital-service-providers, retrieved 2020-12-10.

[42] COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, A new guide regarding security of personal data, https://www.cnil.fr/en/new-guide-regarding-security-personal-data, retrieved 2020-12-10.

[43] PolicyCLOUD, *D3.1 Cloud infrastructure incentives management and data governance: design and open specification 1*, Munné Ricard, 2020.

[44] ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Personal data breach notification under Regulation 2016/679 (wp250rev.01), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612052, retrieved 2020-12-10.

[45] ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Data Protection Officers ('DPOs') (wp243rev.01), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048, retrieved 2020-12-10.

[46] Simkus A. & Van Eecke P. (2020), *European Union: Regulating Big Data – European Commission Introduces Data Governance Bill*, https://www.mondaq.com/uk/data-protection/1010268/regulating-big-data-european-commission-introduces-data-governance-bill (retrieved 2020-11-28).

[47] EUROPEAN DATA PROTECTION BOARD, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, Guidelines 07/2020 on the concepts of controller and processor in the GDPR | European Data Protection Board (europa.eu), retrieved 2020-12-10.

[48] PolicyCLOUD, *D2.1 State of the art & requirements analysis*, Ebro S., 2020.

[49] PolicyCLOUD, *D6.3 Use Case Scenarios Definition & Design*, Sancho Javier, 2020.

[50] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 05/2012 on Cloud Computing*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf, retrieved 2020-12-10.

[51] COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, *Recommendations for companies planning to use Cloud computing services*, Recommendations for companies planning to use Cloud computing services.pdf (cnil.fr), retrieved 2020-12-10.

[52] EUROPEAN UNION AGENCY FOR CYBERSECURITY, *Cloud Computing Risk Assessment*, https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment, retrieved 2020-12-10.

[53] CLOUD SECURITY ALLIANCE, *Cloud Controls Matrix (CCM)*, https://cloudsecurityalliance.org/research/cloud-controls-matrix/, retrieved 2020-12-10.

[54] PolicyCLOUD, *D4.1 Reusable Model & Analytical Tools: Design and Open Specification 1*, Biran Ofer, 2020.

[55] PolicyCLOUD, *D2.2 Conceptual Model & Reference Architecture*, Kiourtis Thanos, 2020.

[56] PolicyCLOUD, *D5.2 Cross-sector Policy Lifecycle Management: Design and Open Specification 1*, Duzha Armend, 2020.

[57] Vacca J.R. (ed.), *Online Terrorist Propaganda, Recruitment, and Radicalization (1)*, CRC Press, Boca Raton.

[58] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Opinion 01/2014 on the "Application of necessity and proportionality concepts and data protection within the law enforcement sector"*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp211_en.pdf, retrieved 2020-12-10.

[59] ARAGÓN OPEN DATA, *Banco de datos*, https://opendata.aragon.es/datos/catalogo?texto=pac, retrieved 2020-12-10.

[60] GOBIERNO DE ARAGON, *SIGPAC. Sistema de información Geográfica de parcelas agrarias*, SIGPAC. Sistema de Información Geográfica de parcelas agrarias. Gobierno de Aragón (aragon.es), retrieved 2020-12-10.

[61] ARAGÓN OPEN DATA, *Términos de uso y licencias*, https://opendata.aragon.es/terminos, retrieved 2020-12-10.

[62] CREATIVE COMMONS, Attribution 4.0 International (CC BY 4.0), Creative Commons — Attribution 4.0 International — CC BY 4.0, retrieved 2020-12-10.

[63] TWITTER, Twitter Terms of Service, https://twitter.com/en/tos, retrieved 2020-12-10.

[64] FACEBOOK, Terms of Service, https://www.facebook.com/terms.php, retrieved 2020-12-10.

[65] FACEBOOK, Automated Data Collection Terms, https://www.facebook.com/apps/site_scraping_tos_terms.php, retrieved 2020-12-10.

[66] LINKEDIN, User Agreement, https://www.linkedin.com/legal/user-agreement, retrieved 2020-12-10.

[67] TWITTER, Twitter Privacy Policy, https://twitter.com/en/privacy, retrieved 2020-12-10.

[68] INFORMATION COMMISSIONER'S OFFICE, What methods can we use to provide privacy information?, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/the-right-to-be-informed/what-methods-can-we-use-to-provide-privacy-information/, retrieved 2020-12-10.

[69] Edwards L. (2016), *Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective*., European Data Protection Law Review, 2016(2), 28–58.

[70] AIRTHINGS, *Terms and Conditions*, https://www.airthings.com/business/terms-and-conditions, retrieved 2020-12-10.

[71] AIRTHINGS, *Data Processing Terms*, https://www.airthings.com/business/data-processing-terms, retrieved 2020-12-10.

[72] NGIoT, *D2.1 End-user engagement and trust building report*, Bolognini Luca, 2020.

[73] PolicyCLOUD, *D1.3 Data Management Plan M12*, Munné Ricard, 2020 (draft available as of 21 December 2020).

[74] Gehra B., Leiendecker J. & Lienke G. (2017), *White Paper. Compliance by Design: Banking's Unmissable Opportunity*, https://image-src.bcg.com/Images/Compliance-by-Design-Dec2017_tcm9-198779.pdf retrieved 2020-11-28.

[75] Balboni P., Taborda Barata M., Botsi A. & Francis K. (2019), Accountability and Enforcement Aspects of the EU General Data Protection Regulation – Methodology for the Creation of an Effective Compliance Framework and a Review of Recent Case Law, Indian Journal of Law and Technology, 15(1), 102-259