CLOUD FOR DATA-DRIVEN POLICY MANAGEMENT

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

D3.9 POLICYCLOUD'S SOCIETAL AND ETHICAL REQUIREMENTS & GUIDELINES – M34

Dissemination Level	PU
Due Date of Deliverable	31/10/2022, M34
Actual Submission Date	31/10/2022
Work Package	WP3 (Cloud Infrastructures Utilization & Data Governance)
Task	3.5
Type	Report
Approval Status	
Version	V1.0
Number of Pages	p.1 – p. 228

Abstract: This report provides a third and definitive version of the identified legal, regulatory, ethical and societal requirements applicable to the Project, as described in Task 3.5. More specifically, this deliverable provides a final report on the identification, refinement, and implementation of all such requirements regarding the PolicyCLOUD platform and each use case, as of the end of the Project (updating the interim implementation report made in its previous version, D3.6).

The information in this document reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at their sole risk and liability.

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
0.1	06/10/2022	First draft	A. Bettiol R. Hanafy I. Oldani M. Taborda Barata
0.2	10/10/2022	Peer Review	R. Munné
0.3	27/10/2022	Address peer review comments	A. Bettiol R. Hanafy I. Oldani M. Taborda Barata
0.4	31/10/2022	Quality Check	A. Mavrogiorgou
1.0	31/10/2022	Deliverable ready for submission	A. Bettiol R. Hanafy I. Oldani M. Taborda Barata

Author List

Organisation	Name
ICTLC	A. Bettiol
ICTLC	R. Hanafy
ICTLC	I. Oldani
ICTLC	M. Taborda Barata

Abbreviations and Acronyms

Abbreviation/Acronym	Definition
API	Application Programming Interface
AI	Artificial Intelligence
CCC	Complementary Customer Control
DMP	Data Marketplace
DPA	Data Processing Agreement
DPIA	Data Protection Impact Assessment
ECI	European Cloud Initiative
EDPB	European Data Protection Board
ENISA	European Union Agency for Cybersecurity
EU	European Union
EUCS	European Union Cybersecurity Certification Scheme for Cloud Services
EULA	End User License Agreement
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding 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
HTTPS	Hypertext Transfer Protocol Secure
IaaS	Infrastructure-as-a-Service
IP	Internet Protocol
IT	Information Technology
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
LIA	Legitimate Interest Assessment
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NER	Named Entity Recognition
OLA	Operational Level Agreement
PDT	Policy Development Toolkit
PM	Policy Model
PME	Policy Model Editor
PP	Public Policy
RDWTI	RAND Database of Worldwide Terrorism Incidents
SDN	Software-Defined Networking
SLA	Service Level Agreement
SMP	Social Media Platform
T&Cs	Terms and Conditions
WP	Work Package
UC	Use Case

Contents

List of Tables

Executive Summary

This report provides the third and definitive version of all the societal and ethical requirements and guidelines as described in Task 3.5. More specifically, this deliverable analyses the ethical, legal, regulatory, and societal issues related to PolicyCLOUD. 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 based on 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, compliance with the requirements defined by the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation – "**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 are 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 data; however, when personal data are involved, appropriate safeguards shall be implemented, especially to comply to applicable data protection laws. Furthermore, this deliverable examines the specific issues related to each of the use cases ("**UCs**"). Finally, a review is provided on how the ethical, legal, regulatory, and societal requirements have been embedded in the solutions developed throughout the Project.

1 Introduction

1.1 Foreword

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

PolicyCLOUD aims to deliver a unique, integrated environment of curated datasets and data management, manipulation, and analysis tools – the PolicyCLOUD platform – addressing the full lifecycle of policy management in distinct thematic areas, and using the data analysis capabilities of the European Cloud Initiative ("**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] [2].

The members of the PolicyCLOUD project (the "**Consortium**") are aware of the necessity of providing extensive and in-depth analyses of legal, regulatory, societal, and ethical aspects, seeing to an optimal embedding of the results of these into the design of the PolicyCLOUD platform, and thoroughly evaluating the extent to which this has been successful, to maximize societal acceptability and trust in the platform and, through this, in policies generated via the platform. Special attention must be paid to the relevant ethical and societal issues. Therefore, 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 have been identified, and a set of requirements, guidelines and norms for the responsible modelling of policies, aligned with the iterations of the development and demonstration of the PolicyCLOUD platform in the relation to the relevant UCs has been proposed in D3.3 [3] and D3.6 [4].

This deliverable, which serves as an update to D3.3 [3] and D3.6 [4], will both report on the progress made in PolicyCLOUD on the refinement and implementation of the set of legal and ethical requirements identified in D3.3 [3] and D3.6 [4], and update those requirements with further legal and ethical issues detected during the course of the project (and after submission of D3.3 [3] and D3.6 [4]), related to previously and newly defined components, as well as to the different UCs.[1] In particular, the main sections of this Deliverable cover the following points:

[1] The London Borough of Camden partner has quit PolicyCLOUD. This partner was piloting one of the UCs. Their specific analytics have been implemented in the remaining UCs to maintain the expected demonstrations. Therefore, no dedicated analysis regarding the UC previously managed by this former partner will be performed in this Deliverable.

- In Section 2, refined sets of legal and ethical requirements, initially derived from D3.3 [3] and D3.6 [4] and subsequently consolidated through discussions with the Consortium, are described for each relevant PolicyCLOUD work package ("**WP**"), as well as steps taken to implement the resulting requirements as of the date of this Deliverable;
- In Annexes Annex 1 – Final WP2 Legal/Ethical Checklist to Annex 23 – WP7 Legal/Ethical Checklist, updated and new lists of ethical and legal requirements are provided, which track the compliance steps implemented to ensure the legal and ethical soundness of the PolicyCLOUD platform. With regards to legal and regulatory issues, the scope of the analysis, in the context of this deliverable, is limited to European Union ("**EU**") and international law, without exploring in detail the specific national and/or local requirements related to the countries and jurisdictions in which the UCs are implemented. More specifically, Annexes Annex 16 – DPIA – UC # 1: Scenario C (Trend Analysis) to Annex 22 – DPIA – UC # 3: Scenario C (Environment and Air Quality Cross-Analysis) provide data protection impact assessment ("**DPIA**") reports regarding each UC using personal data.

1.2 Summary of Changes

This Deliverable aims to provide the final update on the implementation status of the set of legal, regulatory, ethical, and societal requirements identified in D3.3 [3] and D3.6 [4]. It is thus generally focused more on practical considerations than theoretical explanations of relevant legal and ethical issues related to PolicyCLOUD, which have already been extensively addressed in D3.3 [3] – to which readers are generally encouraged to refer for greater detail on the theoretical context surrounding the requirements identified in this Deliverable, as this has not been repeated to avoid redundancy and facilitate its legibility. For the same reason, the analyses and criteria related to the identification of the relevant ethical, legal, regulatory, and societal requirements included in the compliance checklists attached to this Deliverable are described in detail in the context of D3.6 [4] – to which the reader shall refer for more information on this matter.

2 Specification, Development, and Implementation of Identified Requirements

2.1 Requirements, Architecture & Innovation (WP2)

As reported in Section 2.1.1 of D3.6 [4], the **WP2 Legal/Ethical Checklist** (see Annex 1 to D3.6 [4]) was developed as a project-wide list of controls derived from the European Union Cybersecurity Certification Scheme for Cloud Services ("**EUCS**") [5], meant to ensure that an appropriate security posture, including adequate and effective technical and/or organisational security measures, is implemented for the PolicyCLOUD platform.

This Checklist was shared with multiple technical Partners who were identified as potential "owners" for each of the Checklist's controls (i.e., Partners who would be able to provide relevant input as to the status of implementation of a given control for the platform's different components), after an initial assessment and assignment of the different controls conducted by ICTLC. We report the feedback received from each Partner and our relevant conclusions in Section 2.1.1.

As reported in Section 2.1.2 of D3.6 [4], it has been agreed that it is not practically feasible to define a retention period concerning personal data related to PolicyCLOUD users which is shorter than the duration of the project (after which the responsibility for the definition of this period will be transferred to the organisation(s) ultimately responsible for the management of the PolicyCLOUD platform at that stage). Furthermore, the definition of retention periods for personal data contained within data sources processed via the PolicyCLOUD platform is the responsibility of the relevant UC Partner identifying each data source for processing (acting as controller for such personal data, under the GDPR). As a result, the issue of data retention is no longer an active Project-wide issue, and rather one which is to be dealt with at the individual UC level (see Section 2.5.2). For this reason, this issue is no longer reported in this Section.

2.1.1 Data Security

The following Partners were identified as primarily relevant on security matters concerning PolicyCLOUD:

- **LXS**: Leader of Task 2.1 (Requirements Elicitation & State of the Art Analysis).
- **UBI**: Leader of WP3 (Cloud Infrastructures Utilization & Data Governance) and Task 3.6 (Data Governance Model, Protection and Privacy Enforcement).
- **EGI**: Leader of Task 3.1 (Cloud Provisioning of the PolicyCLOUD Infrastructure) and Task 3.2 (Registration of the Cloud-based Environment for Data-driven Policy Management in EOSC).

- **UPRC**: Leader of Task 3.3 (Cloud Gateways & APIs for Efficient Data Utilization), Task 4.2 (Enhanced Interoperability & Data Cleaning) and Task 5.6 (Policy Implementation & Compliance Monitoring); and
- **IBM**: Leader of WP4 (Reusable Models & Analytical Tools), Task 4.1 (Cross-sector Data Fusion Linking) and Task 4.6 (Optimization & Reusability of Analytical Tools).
- **OKS**: Leader of Task 4.5 (Social Dynamics & Behavioural Data Analytics) and Task 5.2 (Modelling & Design of Policies); and
- **MAG**: Leader of WP5 (Cross-sector Policy Lifecycle Management), Task 5.5 (Experimentation, Adaptation & Optimization of Policies and Task 6.3 (Use Cases Implementation & Experimentation).

This identification resulted from an initial assessment and assignment of the EUCS controls included in the **WP2 Legal/Ethical Checklist**, as reported in Annex 1 to D3.6 [4]. Each Partner was engaged to provide feedback on the controls assigned to them (in terms of the accuracy of the assignment and the status of implementation of correctly assigned controls). Based on this feedback, the assignment of controls was revised over time up to the date of this Deliverable.

Please see the **WP2 Legal/Ethical Checklist** (see Annex 1 – Final WP2 Legal/Ethical Checklist) for a record of all input gathered from the abovementioned technical Partners on the status of implementation of each of its controls.

As can be seen from the **WP2 Legal/Ethical Checklist**, the EUCS controls were implemented to a significant extent in PolicyCLOUD. For the majority of controls which were marked as not implemented, this was due to either a lack of (1) technical relevance to PolicyCLOUD, (2) feasibility in their implementation, given the priorities and limited resources available to be allocated to PolicyCLOUD by the relevant Partners, or (3) specific information on how they may have been implemented – however, for most of those controls, the Partners agreed that their implementation could easily be remedied in a potential commercialisation phase for the Platform. It is thus reasonable to maintain, considering the information gathered, that PolicyCLOUD has successfully implemented adequate technical and organisational security measures to ensure secure Platform development (and potential subsequent commercial roll-out).

2.2 Cloud Infrastructure Utilization & Data Governance (WP3)

As reported in Section 2.2 of D3.6 [4], the **WP3 Legal/Ethical Checklist** (see Annex 2 to D3.6 [4]) was developed as a list of controls reflecting a revision and consolidation of the legal, regulatory, ethical and societal requirements deemed relevant to the project's cloud-based infrastructure in D3.3 [3].

The main steps taken to address each of the Checklist's controls were initially reported in Sections 2.2.1 to 2.2.5 of D3.6 [4]. In the below sections, we will describe the additional steps taken to finalize the implementation of all controls as of the date of this Deliverable and provide a final assessment on the implementation of the legal, regulatory, ethical and societal requirements deemed relevant for WP3.

2.2.1 Infrastructure-as-a-Service ("IaaS") Provider Framework

As reported in Section 2.2.1 of D3.6 [4], the following steps had been taken concerning Controls no. 1, 2 and 8 of the **WP3 Legal/Ethical Checklist**, as of the date of that Deliverable:

- A service level agreement ("**SLA**") was entered into between the Consortium and the relevant Partner (EGI) [6], and an operation level agreement ("**OLA**") was entered into between EGI and the PolicyCLOUD platform's IaaS provider [7]. The combination of these documents created several important stipulations pertaining to the provision of cloud-based infrastructure services to PolicyCLOUD, notably a clear definition of services, service hours, relevant exceptions to service hours, response times concerning security incidents and service requests, service level targets, service limitations and constraints, and EGI/the provider's responsibilities.
- Continuous monitoring of the provider's performance by EGI since the execution of the OLA and SLA, to ensure the correct delivery of resources and services requested – a first service performance report was issued with reference to the period between August 2020 to January 2021, and satisfaction reviews have been regularly conducted by EGI with the relevant Partners.
- EGI has made template data processing agreements ("**DPAs**") available, which may be used to complement the OLA and SLA with additional obligations upon EGI/the provider (in particular, the template DPA with EGI Foundation as a Processor relevant to Cloud Computing services) [8].

As of the date of this Deliverable, the mentioned template DPA provided by EGI has been assessed in light of the requirements set out under Article 28 GDPR, which defines the minimum required content of the contract/legal act which should regulate the relationship with a processor (in this case, EGI, with the IaaS provider acting as a sub-processor), and the European Data Protection Board ("**EDPB**") Guidelines 07/2020 on the concepts of controller and processor in the GDPR [9]. The assessment's conclusion was that the DPA met the requirements of Article 28 GDPR overall, although some changes/amendments were recommended to ensure full alignment with these requirements and with best practices indicated by the EDPB.

Despite this favourable assessment, practical limitations were identified regarding the signature of a DPA with the IaaS provider – in particular, the lack of time to complete the negotiation process before the end of the Project (after which the provider would cease to become relevant, as its services would cease) and the difficulties in identifying the correct data protection roles to assign to the counterparty which would sign the DPA with EGI (notably, whether the Consortium should sign the DPA as a whole, or only specific Partners). As a means to provide assurances as to the lawfulness of the personal data processing carried out by EGI/the IaaS provider on behalf of the Consortium despite these limitations, EGI provided a written confirmation of its role as processor (and, implicitly, of the IaaS provider as sub-processor) regarding the personal data used and processed by the Consortium on the resources provided by EGI to PolicyCLOUD (i.e., the provision of resources framed by the OLA and SLA mentioned above). They further confirmed that all such processing took place in accordance with the terms of the mentioned template DPA, EGI's relevant policies [10] - notably their Policy on the Processing of Personal Data [11] – and a description of technical and organisational security measures implemented by EGI to meet legal and contractual requirements when processing personal data [12].

As this confirms that the processing of personal data conducted by EGI and the IaaS provider within the context of PolicyCLOUD has abided by Art. 28 GDPR, it is reasonable to maintain that such a confirmation serves as an acceptable, albeit sub-optimal (from a compliance perspective) alternative to formally closing out a DPA for PolicyCLOUD. As such, this Control is now considered as fully implemented, as described in the final **WP3 Legal/Ethical Checklist** (see Annex 2 – Final WP3 Legal/Ethical Checklist).

2.2.2 Environmental Impact

As reported in Section 2.2.2 of D3.6 [4], the only step which had been taken concerning Control no. 3 of the **WP3 Legal/Ethical Checklist**, as of the date of that Deliverable, was the identification of a relevant owner for its implementation (i.e., EGI, which should engage the IaaS provider for further information).

As of the date of this Deliverable, EGI has obtained confirmation from the IaaS provider that:

- They closely monitor the power usage effectiveness of the data centre which supports the PolicyCLOUD infrastructure (to reduce inefficiencies which might cause unnecessary energy consumption, which in turn mitigates the data centre's environmental impact).
- Energy consumption optimisation measures are in place, notably the ability to tune the cooling facilities used and adjust the amount of IT resources which are active at any given time.
- The implementation of a direct free-cooling solution, which could rely on natural fresh air of up to 23º C in temperature to regulate cooling (and which could provide greater power usage effectiveness than mechanical cooling), is planned.

This indicates that technical and/or organisational measures to ensure that the environmental impact of the PolicyCLOUD project's cloud-based infrastructure is reduced to a minimum have been implemented. As such, this Control is now considered as fully implemented, as described in the final **WP3 Legal/Ethical Checklist** (see Annex 2 – Final WP3 Legal/Ethical Checklist).

2.2.3 Data Security

As reported in Section 2.2.3 of D3.6 [4], data security concerns related to PolicyCLOUD – including those relevant to its cloud-based infrastructure – are more generally covered by the **WP2 Legal/Ethical Checklist** (see Section 2.1). This also includes requirements regarding contractual limitations on individuals' ability to process personal data collected and managed via the PolicyCLOUD platform, with reference to Consortium personnel (as PolicyCLOUD users' limitations are addressed by the relevant terms and conditions ("**T&Cs**") – see Section 2.4 for more information).

Furthermore, as also reported in that Section of D3.6 [4], a Data Governance and Privacy Enforcement mechanism has been developed for the PolicyCLOUD platform, through which it is possible to determine whether PolicyCLOUD users are able to access a given piece of data, file, database, folder or other "object" hosted on the platform's cloud-based infrastructure, based on several contextual attributes. The

Platform itself has further been crafted in such a manner as to support various data governance models. For more on this mechanism, please refer to Section 6 of D3.7 [13].

This indicates that technical and/or organisational measures to ensure the security of data hosted on the PolicyCLOUD's cloud-based infrastructure, as a result of a dedicated security risk assessment (and meeting the goals described in Section 2.2.3 of D3.6 [4]), have been implemented. As such, this Control is now considered as fully implemented, as described in the final **WP3 Legal/Ethical Checklist** (see Annex 2 – Final WP3 Legal/Ethical Checklist).

2.2.4 Data Subject Rights

As reported in Section 2.2.4 of D3.6 [4], it is necessary to ensure that the cloud infrastructure on which the Platform is hosted allows for the exercise of rights granted by the GDPR, or by other applicable laws, to individuals whose personal data may be processed via the platform ("**Data Subject Rights**") – at a minimum, this infrastructure should not create any relevant technical obstacles to the exercise of these rights.

It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the Platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered, regarding users of the PolicyCLOUD platform, via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the Platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of Data Subject Rights, concerning PolicyCLOUD users (for other relevant individuals, e.g., individuals whose personal data may be contained in data sources processed via the Platform, see Section 2.3.4).

This indicates that technical and/or organisational measures to allow for the exercise of Data Subject Rights to individuals whose personal data may be processed on the Platform, concerning PolicyCLOUD users, have been implemented. As such, this Control is now considered as fully implemented, as described in the final **WP3 Legal/Ethical Checklist** (see Annex 2 – Final WP3 Legal/Ethical Checklist).

2.2.5 Data Retention

As reported in Section 2.2.5 of D3.6 [4], it is necessary to ensure that personal data is not stored on the Platform for any longer than necessary to allow the purposes for its lawful collection and further processing to be met.

As noted in Section 2.1, it has been agreed that it is not practically feasible to define a retention period concerning personal data related to PolicyCLOUD users shorter than the duration of the project (after which the responsibility for the definition of this period will be transferred to the organisation(s) ultimately responsible for the management of the Platform at that stage). As further reported in Section 2.2.5 of D3.6 [4], it has been confirmed that a mechanism for deletion of such personal data (regarding

PolicyCLOUD users – for other relevant individuals, e.g., individuals whose personal data may be contained in data sources processed via the PolicyCLOUD platform, see Section 2.3.5) has been implemented within the Platform, and that the cloud-based infrastructure on which the Platform is hosted does not present any relevant technical obstacles to the implementation of this ability and, consequently, to the possibility to enforce defined retention periods for such personal data.

This indicates that technical and/or organisational measures to ensure that, after a defined retention period is exceeded, personal data on PolicyCLOUD users can be effectively erased or anonymised, have been implemented. As such, this Control is now considered as fully implemented, as described in the final **WP3 Legal/Ethical Checklist** (see Annex 2 – Final WP3 Legal/Ethical Checklist).

2.2.6 Incentives Management

As reported in Section 5.1 of D3.7 [13], PolicyCLOUD's reviewers have indicated, in their recommendations, that it would be appropriate not to continue with the development of the Incentives Management component, and to redirect the efforts initially planned for this task to the improvement of the policy development toolkit ("**PDT**"). As such, there are no updates to be provided regarding this component in this Deliverable. We refer to Section 3.1.2 of D3.6 [4], which includes an assessment of applicable legal, regulatory, ethical and societal requirements based on the information available on the Incentives Management tool as of the date of that Deliverable, and which would remain relevant now if the development of this component were to continue.

2.3 Reusable Models & Analytical Tools (WP4)

As reported in Section 2.3 of D3.6 [4], the **WP4 Legal/Ethical Checklist** (see Annex 3 to D3.6 [4]) was developed as list of controls reflecting a revision and consolidation of the legal, regulatory, ethical and societal requirements deemed relevant to PolicyCLOUD's data transformation and analytics components in D3.3 [3], as well as the possibility to register additional analytics tools (usable in the transformation/"pre-processing", interoperability enhancement and/or structured output analysis phases described above) and data sources on which such tools (whether pre-existing on the platform, or subsequently registered) can be applied.

The main steps taken to address each of the Checklist's controls were initially reported in Sections 2.3.1 to 2.3.5 of D3.6 [4]. In the below sections, we will describe the additional steps taken to finalize the implementation of all controls as of the date of this Deliverable and provide a final assessment on the implementation of the legal, regulatory, ethical, and societal requirements deemed relevant for WP4.

2.3.1 Analytics Compliance (including Analytics Tool Registration)

As reported in Section 2.3.1 of D3.6 [4], the following steps had been taken concerning Controls no. 1, 3, 4 and 10 of the **WP4 Legal/Ethical Checklist**, as of the date of that Deliverable:

- It was decided, in the context of WP4, that the specific Consortium members responsible for each relevant tool would be engaged to provide information about the training and testing protocols/programs, mechanisms facilitating auditability of AI-based systems (including the traceability of the development process, the sourcing of training data used, the logging of processes/outcomes/impacts), the trade-offs between applicable legal/ethical requirements and principles considered during design, and the degree of possibility of false positive/negative correlations, applicable (or to be applied, where relevant) to the tools for which they are responsible.
- Additional input parameters (*biasDoc* and *tradeoffsDoc*) were added to the analytics function registration Application Programming Interfaces ("**APIs**"), requiring registrants to link bias and trade-off management information/documentation to their tools upon registration (see Sections 2.1.2 and 2.2.3 of D4.5 [14]).

To help to ensure that PolicyCLOUD users are advised of the possibility that false positive or negative correlations arise because of the use of analytic functions registered on the Platform, so that those users are incentivised to verify the validity of the correlations made and results presented to them on the platform, a warning has been added to the PDT and policy model editor ("**PME**") User Handbook (see Section 2.4 for more information). Furthermore, a specific e-mail address will be set up on the PDT/PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during Platform use (see Section 2.4 for more information). In the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the Platform) may be implemented, if deemed feasible and effective.

As subsequently reported in Section 2.3.3 of D4.5 [14], two lists of questions were developed to serve as a guide for analytics function registrants ("**Analytic Owners**") to understand the type of information which should be provided to address these two additional input parameters, in a manner which is complete and effective towards future Platform users (considering that the information provided will be linked to the registered functions for those users' benefit). These questions were included in a brief Registration User Guide, as part of the guidance provided to Analytic Owners on addressing each of the registration procedure's input parameters (see Annex 26).

These questions were also presented to the Partners responsible for the Platform's built-in analytic functions, to document the approaches taken on bias and trade-off management for such functions. After discussion with those Partners, conclusions were arrived at which indicate that appropriate bias and trade-off management measures have been implemented for each of these functions. As such, this Control is now considered as fully implemented, as described in the final **WP4 Legal/Ethical Checklist** (see Annex 3 – Final WP4 Legal/Ethical Checklist).

The main conclusions arrived at for each function are reflected below:

2.3.1.1 Data Cleaning

Concerning **bias management**:

- The Data Cleaning "analytic function" should be seen more as a data pre-processing tool than an analytics function *per se*. Rather than analyze a dataset and derive relevant conclusions or insight from such input, the goal of Data Cleaning is to ensure that all data collected from potentially heterogeneous data sources is as clean and complete as possible, notably by detecting and correcting or removing inaccurate or incomplete datasets, and by replacing, modifying or deleting incorrect, inaccurate or irrelevant data within specific datasets;
- Having said this, the definition of adequacy constraints and rules which are used by Data Cleaning to filter out inappropriate data is an ethically-relevant choice – any biases or prejudices inherent to constraint definition will potentially apply to the resulting "cleaned" dataset. As such, constraint definition should be guided by objective and reasoned criteria, seeking only to ensure the relevance of the content left within cleaned datasets for the purposes of the specific project in which Data Cleaning is used – this requires, for example, each of the PolicyCLOUD's UC Partners to carefully consider which datapoints are relevant to their specific UC scenarios, for each of the data sources they identify as relevant to PolicyCLOUD. This, however, is an issue of potential bias in the use of Data Cleaning in a specific context and by a specific user, rather than an issue of inherent bias in the functioning of Data Cleaning itself. In the context of PolicyCLOUD, the UC Partners have been advised of this and have been supported in the management of their data sources through legal/ethical assessments provided by ICTLC (see Section 2.5 for more on this). For hypothetical future users, a disclaimer can be added (e.g., in the context of a user guide, on the PDT itself, or in a user interface which may be designed for users to input Data Cleaning rules and constraints) to advise such users of the need to appropriately consider the risk of bias in data source selection and constraint definition (see Section 2.3.1, above, for more on this).

Concerning **trade-off management**:

- Data Cleaning can be used as a tool to protect interests and values such as personal data privacy, if it is set up by users appropriately (i.e., by defining constraints in line with the GDPR's data minimization principle).
- Given that Data Cleaning should not be regarded as an analytic functions *per se*, the issue of actual and/or potential trade-offs was not explored in-depth prior to the Project. At any rate, trade-offs arising in relation to Data Cleaning must do more with the definition of constraints than the design of the component itself. For example:
    - <u>Privacy and Security vs. Accuracy.</u> In general, the larger the amount of data included in a dataset (excluding corrupted or inaccurate data), the more accurate the result of a subsequent analytic function applied to that dataset will be. However, where personal data are included, increasing the amount of data also increases the possibility of a breach of the GDPR's data minimization principle (if unnecessary or irrelevant personal data are

collected and kept beyond the cleaning process) and the degree of risk to the rights, freedoms, and interests of data subjects – particularly in the event of a personal data breach affecting the stored data.

- <u>Accuracy vs. Fairness.</u> The definition of constraints may seek to exclude certain datapoints to help prevent discriminatory outcomes (e.g., in a dataset containing personal data, removing information identifying an individual's racial origin, directly or indirectly). However, depending on the datapoints excluded, the accuracy of subsequent analytics operations may be inadvertently lessened (e.g., where addresses are excluded from a dataset because they may indirectly suggest an individual's racial origin, leading to a lessened ability for the analytics operations to determine the geographic distribution of the population within that dataset).

2.3.1.2 Enhanced Interoperability

Concerning **bias management**:

- Enhanced Interoperability seeks to promote accuracy of the ultimate results reached by the analytic process, in that it creates connections between heterogeneous data sets which should allow subsequent analytic functions to operate on them jointly (and thus expectedly increase the accuracy of the analytic output, given the wider base of data used as input).
- The evaluation and assessment of the tool was performed based on its utilization on the datasets of UC#1, and especially on the Global Terrorism Database ("**GTD**"), where data related with specific groups of people were analyzed and processed (see Sections 2.5.1.1 and 2.5.2.1 for more on this). In this respect, no issues and indications of unfair or biased results were reported, and thus the tool was finally registered in the Platform.
- Researchers claim to have found evidence of demographic bias in the NER task, which is part of Enhanced Interoperability – in particular, that NER models using synthetically-generated data are generally better at identifying names of a given racial origin than those of others, and that even techniques such as debiased embeddings may not suffice to resolve such bias [37]. However, there is no UC scenario under PolicyCLOUD relying on Enhanced Interoperability using data which might trigger such a risk of bias.
- NER is a subtask of natural language processing and Artificial Intelligence ("**AI**") in general. In the PolicyCLOUD context, the widely used open-source library spaCy 2.0 [38] was used to implement this task. spaCy v2.0's NER system offers different pre-trained NER models which are based on the use of, trained on an OntoNotes dataset [39]. In particular, the "spacy_lg" model was used, which relies on global word vectors (GloVe) to perform its word embedding strategy, and which has shown good accuracy and low bias in relevant bias-related studies [37].
- The corpus used to develop and train Enhanced Interoperability is public, but no actions have been performed to check this corpus for specific biases. Research has proven that gender bias may exist within this corpus [40] but, for the purposes of PolicyCLOUD, no such bias has been identified.

- Further research should be conducted (e.g., in a potential commercialization phase) to identify potential biases which may arise from the application of NER to the examined PolicyCLOUD datasets, as well as to assess Enhanced Interoperability against the AI Fairness 360 Toolkit (or alternative tools).
- Given that Enhanced Interoperability processes and identifies entities, the results of which are made available to other analytic functions to provide a final output, all limitations in terms of bias which can be linked to Enhanced Interoperability are more directly connected to those other analytic functions. Without prejudice to this, Enhanced Interoperability further mitigates the risk of bias in entity identification by relying also on semantic information derived from ontologies and taxonomies (provided by the relevant UC Partner or a given data provider) to boost the accuracy of its results.

Concerning **trade-off management**:

- Enhanced Interoperability should be seen rather as a data pre-processing tools (meant to prepare data for further analysis by analytic functions) than an analytic functions *per se*, although Enhanced Interoperability is based on an AI model. This issue was therefore not explored in-depth prior to the Project.
- However, in general, trade-offs related to Enhanced Interoperability can be analyzed to conclude that they have to do with the definition of constraints, as well as with the design of the component itself. For example:
    - Explainability vs. Accuracy. The specific actions taken by Enhanced Interoperability to produce its output follow an interpretable and explainable approach to entity recognition, allowing users to understand how entities are recognized within a given input (e.g., text). Enhanced Interoperability further makes use of semantic information provided by the data provider itself, which creates a high degree of correlation between the quality of the output (and resulting semantic interoperability) and the degree of information made available by that data provider. This may, overall, impact the overall accuracy of the spaCy model used to a limited extent – this is a point which may be further evaluated (e.g., in a potential commercialization stage) in collaboration with the relevant Partners, considering also the final analytical output generated by their analytic functions (which rely on Enhanced Interoperability).
    - Accuracy vs. Fairness and Privacy. The definition and provision of semantic information from data providers may help to eliminate biases and unfairness in the output generated by Enhanced Interoperability, and to enhance data protection by design (where personal data may be included in the input). However, where this implies excluding information from the input, the accuracy of the output generated by Enhanced Interoperability (and, subsequently, by the other analytic functions relying on its output) may be lessened.

2.3.1.3 Sentiment Analysis

Concerning **bias management**:

- Sentiment Analysis' development is based on the usage of the VADER [34] or BERT [35] models. This is potentially subject to change in the future (e.g., in a potential commercialization phase), as the use of other sentiment analysis libraries is currently being considered. It has been trained using a corpus of data sets, which is available for inspection.
- The Sentiment Analysis analytic function makes use of external sentiment analysis libraries, which commonly involve some degree of bias resulting from the system training methodology used. For example, in the creation process for the libraries used in Sentiment Analysis, aspects which may lead to biased results include: (1) the inclusion of western-style emoticons within the libraries which, when applied to non-western-style texts, may generate wrong results; and (2) the use of ten human raters in the creation of manual data labels used within the libraries – as no information is available on the age, gender, class or nationality (*inter alia*) of these raters, it is not possible to exclude that the group of raters may not be sufficiently representative of the population which is analyzed via Sentiment Analysis (which may lead to wrong results based on the criteria applied by the raters in generating labels).
- Due to the complex nature of the analytics operations performed by Sentiment Analysis, it is not possible to develop a system to detect and address all sources of bias arising from the use of external libraries from scratch. Such external libraries typically also do not consider all biases which may arise in the natural language processing field.
- It is considered impossible to guarantee fair and bias-free results in analytics functions such as Sentiment Analysis (an issue that has arisen also in other state-of-the-art approaches to this matter) [36] . There are several factors that contribute to this, such as (1) the subjective nature of the predicted attributes (as Sentiment Analysis aims at predicting an affective representation value); (2) the common methods of development of analytics functions such as Sentiment Analysis (as most are based on one or several labeled datasets, which usually do not cover a wide range of cultural, racial and gender backgrounds in their training process); and (3) the prohibitive costs that developing a proper, bias-free, focused validation of an analytics function such as Sentiment Analysis would mean (as some functions require human validation, a proper bias-free focused validation would require involving experts with strong expertise in gender issues, as well as in all racial and cultural backgrounds, capable of assessing all the different potential ways in which these aspects might have an impact on written text).
- Reliance on Sentiment Analysis thus requires an understanding of these potential biases to ensure that results generated by this analytic function are assessed critically, rather than merely taken at face value. By documenting these considerations and making them available in a user-friendly manner, conclusions reached by Sentiment Analysis users can be interpreted with these limitations in mind.

Concerning **trade-off management**:

- Although Sentiment Analysis is an analytical function, the subjectivity of the analysis it is meant to perform makes it impossible to fully assess the degree to which different interests and values are given preference over others without considering the specific data sources it is to be applied to in the Project. As such, this assessment could not be conducted a priori, but must rather be considered subsequently to the selection of the data sources which are to be analyzed (considering the types and content of the data to be analyzed, as well as the type of output to be produced).
- Considering the UC#2 relevant data sources, there is one main trade-off which can be highlighted as relevant at this stage:
    - Privacy vs. Accuracy and Explainability. Sentiment Analysis will provide a more accurate overview of the sentiment regarding a given topic if the data source to which it is applied is larger. However, increasing the amount of data subjected to analysis may also increase privacy risks for individuals, particularly where this data amounts to social media posts (made by individual and identifiable users). To allow for more accurate results without compromising the privacy and data protection rights of social media users, ingested data to be assessed via Sentiment Analysis should be purged of relevant identifiers (such as names and usernames, which are not needed in any case for analysis to be conducted). Additionally, Sentiment Analysis' output should be complemented with metadata about the analytic operations performed, to better contextualize such output and ensure that PolicyCLOUD users are able to properly interpret results (thus mitigating the risk of misinterpretation and subsequent misguided policymaking activities) – this information should include, for example, statistics on the number and geographical provenance of social media posts analyzed to produce a given output, an explanation of the process followed by Sentiment Analysis to produce that output, a disclaimer warning users to exercise their own critical judgment in interpreting results. To go even further would include aggregated details on the social media users themselves (e.g., gender, nationality, location percentages), though this would increase the privacy risk to those data subjects further – this should, as such, not be considered unless there is a strict need to further contextualize Sentiment Analysis' output to render it effectively useable for policymaking purposes.

2.3.1.4 Variable Distributions

Concerning **bias management**:

- The Variable Distributions analytic function is conceived as an exploratory data analysis tool in charge of summarizing and visualizing categorical variables (frequency distributions, spatial distribution, etc.). It relies on an aggregation process, which is performed over one or two such variables selected by the user, to produce its output.
- Variable Distributions does not analyse a dataset and derive relevant conclusions or insight from such input. As such, any risk of bias in the output provided by Variable Distributions is not

inherent to the function in itself, but rather dependent on the data source to which it is applied (as any biases inherent to that data source may be reflected in this output) – where this data source is biased (e.g., because the data collection method used does not represent a target population accurately, due to the use of biased sources or the collection of insufficient datapoints), Variable Distributions will not be able to mitigate this bias and may produce skewed results.

- In the context of PolicyCLOUD, the UC Partners have been advised of this and have been supported in the management of their data sources through legal and ethical assessments provided by ICTLC (see Section 2.5 for more on this). For hypothetical future users, a warning has been added to the PDT-PME Handbook to advise such users of the need to appropriately consider the risk of bias in data source selection and variable selection (as noted in Section 2.3.1).

Concerning **trade-off management**:

- Given that Variable Distributions should not be regarded as an analytic function per se (see above), the issue of actual and/or potential trade-offs was not explored in-depth prior to the Project.
- At any rate, trade-offs arising in relation to Variable Distributions must do more with the selection of variables than the design of the component itself. For example:
    - <u>Privacy and Security vs. Accuracy.</u> In general, the larger the amount of data included in a dataset to which Variable Distributions is applied, the more accurate the results provided by Variable Distributions will be (assuming appropriate variables have been selected). However, where personal data are included, increasing the amount of data also increases the possibility of a breach of the GDPR's data minimisation principle (if unnecessary or irrelevant personal data are collected and kept beyond the cleaning process) and the degree of risk to the rights, freedoms and interests of data subjects – particularly in the event of a personal data breach affecting the stored data.
    - <u>Accuracy vs. Fairness.</u> The selection of variables may seek to exclude certain variables to help prevent discriminatory outcomes (e.g., in a dataset containing personal data, removing variables which address individuals' racial origin, directly or indirectly). However, depending on the variables excluded, the results which might be inferred from the output provided by Variable Distributions may be inadvertently skewed (e.g., where address is excluded as a variable because it may indirectly suggest an individual's racial origin, leading to a lessened ability to determine the geographic distribution of a population within that dataset).

2.3.1.5 Time Series Forecasting

Concerning **bias management**:

- Time Series Forecasting is a machine learning ("**ML**")-based tool in charge of forecasting future data based on a historical dataset. In the PolicyCLOUD context, several Python libraries have been used to apply the models generated via Time Series Forecasting, including the SARIMA model [30] and Prophet for Univariate Time Series Forecasting ("**Prophet**") [31]. Further research work should be conducted to identify potential biases inherent to the application of these models to PolicyCLOUD datasets.

- Before applying the Time Series Forecasting analytic function to any given UC, it is important to conduct a preliminary identification of potential dataset bias (regarding the dataset(s) relevant to that use case). Such biases may arise during distinct stages of the workflow involving the analytic function, including data collection, dataset pre-processing, model generation, results presentation, output analysis, policymaking activities, etc.
- Further bias management steps may be taken (potentially at a commercialisation stage), such as: the inspection of the training corpus/datasets used to develop Time Series Forecasting (which is available for this purpose), the identification and application of tools for model result analysis (e.g., IBM's AI Fairness 360 Toolkit), and the development of guidelines on potential bias identification during the data processing stages of PolicyCLOUD.
- Regarding performance variance bias (e.g., where Time Series Forecasting is unable to capture all the elements in a given time series, and consequently provides a high number of errors during its training and, subsequently, the prediction and output phase), a multiple-time-series-predictors approach has been used to monitor and review the results produced by the function for possible biases. In particular, the SARIMA model [30] has been selected, fitted, and trained to review Prophet [31] (and, therefore, Time Series Forecasting) forecast accuracy.
- Since the analytical process to be conducted via Time Series Forecasting in the context of PolicyCLOUD only considers one variable, and the output generated is a prediction of the value of that same variable, the results which may be generated by Time Series Forecasting cannot be biased by data pertaining to other variables, which are not considered. In short, given the simplicity of the context in which Time Series Forecasting is to be deployed in the context of PolicyCLOUD, it was deemed disproportionate to run a structured bias testing procedure on this function – however, this is a step which will be taken in a potential commercialisation phase for the Platform.

Concerning **trade-off management**:

- The accuracy of results and some other performance indicators were identified as relevant interest and values to the functioning of Time Series Forecasting. As a ML-based prediction tool, evaluating the ML model's performance is a key aspect in its functioning. Different accuracy metrics have been defined for this performance evaluation, such as: mean squared error, mean absolute error ("**MAE**") and Root Mean Squared Error, with MAE being the metric of reference, as it is more robust to outliers.
- To avoid concerns around personal data privacy, the data sources which are being used in the design, development and implementation of Time Series Forecasting in the context of PolicyCLOUD are aggregated data entirely based on a temporal variable, which therefore cannot be used to identify specific individuals.

- Given that no use of personal data, nor risk of exclusion of relevant variables, has been identified in the context of PolicyCLOUD, the issue of actual and/or potential trade-offs was not explored in-depth. More on this follows:
  - Privacy and Security vs. Accuracy. In general, the larger and newly the amount of data included in a dataset for training the models, the more accurate the results provided by prediction will be. However, since no personal data were included in the training dataset used for Time Series Forecasting or in the datasets to which it is applied, the possibility of a breach of the GDPR's principles and the degree of risk to the rights, freedoms and interests of data subjects is not considered relevant.
  - Accuracy vs. Fairness. The dataset used in Time Series Forecasting's training phase was generated by aggregating original dataset rows using exclusively a temporal variable. As such, no discriminatory outcomes which could arise due to errors in the elimination of critical variables are expected to arise from the use of Time Series Forecasting in the context of PolicyCLOUD. Since all variables, except a temporal variable, are excluded, there is a low probability of inferring skewed results from Time Series Forecasting's output. As for fairness, Time Series Forecasting relies mostly on commonly used machine learning algorithms, for which no issues regarding fairness have been identified in currently available literature.
  - Explainability. Certain third-party libraries (Prophet [32] and Statsmodel [33]) used to develop and implement Time Series Forecasting provide some functionalities in terms of explainability. In both cases, outcomes generated by Time Series Forecasting using these libraries are enriched by a set of confidence intervals which may help to explain those outcomes; however, for the sake of readability and management of limited visualization resources, no additional traceability information is provided to PolicyCLOUD users regarding these outcomes (though it would be possible to amend this). Additionally, Time Series Forecasting's output should be complemented with metadata about the analytic operations performed, to better contextualize this output and ensure that PolicyCLOUD users are able to properly interpret results (thus mitigating the risk of misinterpretation and subsequent misguided policymaking activities) – steps which can be contemplated in a potential commercialisation phase would be to provide such information in the PDT-PME, including, for example, an explanation of the process followed by its models to produce its output, a disclaimer warning users to exercise their own critical judgment in interpreting results (which is included in the latest draft of the PDT-PME Handbook), etc.

2.3.1.6 Social Dynamics

Concerning **bias management**:

- The Social Dynamics analytic function was subjected to an assessment of its limitations in achieving fair and unbiased results prior to its registration on the Platform, with no cases of generation or reinforcement of potential or unfair biases caused by the function detected.

- Social Dynamics executes simulation models with completely transparent assumptions, mechanics and outcomes (given that they are described in code, which is fully visible to the function user). As a result, Social Dynamics does not itself carry a relevant risk of bias in its design; however, there may be a risk of bias inherent to the submitted models which are executed via Social Dynamics.
- As a relevant bias mitigation measure concerning the risk of bias which may arise from a fixed execution order established for the dynamic rules set up for each individual node within a given model, Social Dynamics' general concurrent algorithm for simulation execution applies random permutations to this execution order.
- Social Dynamics is not based on an AI model.
- Social Dynamics allows its users to clone simulation models and execute each clone with different social network models, modelling assumptions and/or mechanics to determine whether such simulation models may be biased. Furthermore, Social Dynamics has a graphical user interface which allows users to define criteria by which to evaluate and compare model outcomes against available bias benchmarks.

Concerning **trade-off management**:

- Social Dynamics is a simulation environment that operates on synthetic data, to achieve generality in its results. This means that it automatically generates a social network based on macroscopic statistical features of a population of interest, which in that describe those macroscopic features in relation to a social group. In other words, the synthetic data used is representative of families of real datasets, but not equal to any of them.
- To produce effective results, synthetic data must accurately reflect the real datasets for which it acts as a replacement – otherwise, it will not reflect the patterns contained in such datasets which are crucial for an accurate analysis outcome. It is also possible for synthetic data to contain inherent or historical bias, where the original underlying datasets are themselves biased. Result accuracy is thus a key concern of Social Dynamics.
- However, in the interest of generating accurate synthetic data, one must not forget the need to protect the fundamental rights of individuals, notably the right to privacy and data protection. Where a synthetic dataset is too similar to one or more underlying real datasets containing personal data, this may create privacy issues to the extent that information within the synthetic dataset can be traced back to identifiable individuals. Personal data privacy therefore also plays a role in synthetic data use (and, in fact, is a key argument in favour of synthetic data use over real data use).
- There are specific trade-offs which can be flagged as relevant to Social Dynamics:
    - Privacy vs. Accuracy. As noted above, the greater the similarities between the synthetic dataset and the underlying real datasets, the greater the likelihood that the synthetic dataset will accurately reflect the patterns contained in the real datasets. This, in turn, ensures greater accuracy in the results of analyses conducted by Social Dynamics using such synthetic data. However, if the synthetic dataset is too similar to the underlying real datasets (containing personal data), it may reveal personal information which can be

traced back to identifiable individuals, thus potentially creating a risk to their right to privacy and data protection. However, in the context of PolicyCLOUD, the construction of synthetic datasets relied on by Social Dynamics does not require an underlying real dataset; instead, synthetic datasets are created based on user-defined general specifications regarding network structure and high-level statistical characteristics. Any similarities shared between the synthetic datasets and real datasets is therefore highly unlikely, and most probably coincidental. The only other way such similarities could arise would be if a user would upload a dataset to Social Dynamics (which has not been generated by Social Dynamics) which is identical or a close replica to a real dataset, in which case the user would need to assume responsibility for any potential processing of personal data which might arise (or otherwise provide assurances of the adequate anonymization of such data).

- Accuracy vs. Fairness. The construction of the synthetic dataset may seek to adjust the underlying real datasets to mitigate the risk of bias creeping into the synthetic dataset, and/or to exclude certain datapoints to help prevent discriminatory outcomes (e.g., in a dataset containing personal data, removing information which identifies an individual's racial origin, directly or indirectly). However, depending on the adjustments made and datapoints excluded, the accuracy of subsequent analytics operations may be inadvertently lessened (e.g., where addresses are excluded from a dataset because they may indirectly suggest an individual's racial origin, leading to a lessened ability for the analytics operations to determine the geographic distribution of the population within that dataset). However, as noted previously, the synthetic datasets generated in the context of PolicyCLOUD do not use real datasets as input. The user-defined specifications which are used instead are explicitly described to users, which should consider them carefully to mitigate the risk of bias or inaccuracy with regards to the population they wish to target with a given policy.

2.3.1.7 Trend Analysis

Concerning **bias management**:

- Though Trend Analysis is not an analytic function based on an AI model, it relies on the output of the Enhanced Interoperability component (see Section 2.3.1.2), as well as on an ontology defining a domain. Mislabeling related to the former and/or errors in the definition of the latter may therefore bias the results reached by Trend Analysis (reflecting or reinforcing the pre-existing biases within the mentioned component/ontology). It is therefore paramount to implement relevant bias and trade-off mitigation measures regarding the former (as described in the mentioned Section), and to thoroughly assess the suitability and risk of bias within ontologies (considering the policymaking goals for which they have been defined, and the data sources to which they are to be applied), to mitigate this risk to the greatest extent feasible.

Concerning **trade-off management**:

- Although Trend Analysis is an analytical function, the extent to which relevant trade-offs may arise depend on the specific data sources to which it is to be applied. As such, this assessment could not be conducted a priori, but must be considered subsequently to the selection of the data sources to be analysed (considering the types and content of the data to be analysed, as well as the type of output to be produced).
- In general, main trade-off which can be mentioned as potentially relevant is the following:
  - Privacy vs. Accuracy. Trend Analysis will provide an overview of stronger trends regarding a given topic if the data source to which it is applied is larger. However, increasing the amount of data subjected to analysis may also increase privacy risks for individuals, particularly where this data amounts to social media posts (made by individual and identifiable users). To avoid compromising the privacy and data protection rights of social media users, ingested data to be assessed via Trend Analysis do not include relevant identifiers (such as usernames) of the authors of such social media posts, which are not needed in any case for analysis to be conducted. Trend Analysis' output is, in any case, aggregated, such that it is not possible to directly connect information within this output to any given individuals.
- Considering the data sources relevant to UC#1, the above issue is made more relevant in light of the fact that the ontology used may contain personal information on identified/potential terrorists (e.g., names and surnames) – this creates a risk not only of non-compliance with the privacy and data protection rights of those individuals under EU law, but also of possible misidentification where that information is relatively common (potentially causing someone to be erroneously identified as a terrorist by Trend Analysis). These risks have, in any case, been mitigated by the measures taken in connection with the legal/ethical assessment of the data sources relevant to UC#1.

2.3.2 Data Security

As reported in Section 2.2.3 of D3.6 [4], data security concerns related to PolicyCLOUD – including those relevant to the analytics tools and data sources hosted on the Platform – are more generally covered by the **WP2 Legal/Ethical Checklist** (see Section 2.1). This also includes requirements regarding contractual limitations on individuals' ability to process personal data collected and managed via the Platform, with reference to Consortium personnel (as PolicyCLOUD users' limitations are addressed by the relevant T&Cs – see Section 2.4 for more information). Such requirements are further complemented, as far as restrictions on the unlawful reuse of personal data are concerned, by the Data Governance and Privacy Enforcement mechanism developed for the Platform (addressed in Section 2.2.3 and in further detail in Section 6 of D3.7 [13]).

Furthermore, as also reported in D3.6 [4], a logging service has been developed for the Platform, through which it is possible to, *inter alia*, identify security threats, analyse suspected incidents (i.e., forensic analysis), monitor internal policy violations, collect information on abnormal events and debug both performance and functional issues, making this critical to ensure the security of the Platform. See Section 2.3.4 of D4.5 [14] for more on this service.

This indicates that technical and/or organisational measures to ensure the security of analytics tools and data sources hosted on the Platform, meeting the goals described in Section 2.2.3 of D3.6 [4], have been implemented. As such, this Control is now fully implemented, as described in the final **WP4 Legal/Ethical Checklist** (see Annex 3 – Final WP4 Legal/Ethical Checklist).

2.3.3 Data Source Registration

As reported in Section 2.3.3 of D3.6 [4], concerning Control no. 5 of the **WP4 Legal/Ethical Checklist**, additional input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) were added to the data source registration APIs, requiring registrants to link bias, and privacy and data protection management information and documentation, as well as information and documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders when needed, to their data sources upon registration (see Sections 2.1.2 and 2.2.2 of D4.5 [14]).

Notes on addressing these input parameters in a complete and effective manner towards future platform users (considering that the information provided will be linked to the registered data sources for those users' benefit) was included in a brief Registration User Guide, as part of the guidance provided to registrants on addressing each of the registration procedure's input parameters (see Annex 26).

The data sources identified by the UC Partners for use within PolicyCLOUD have been subjected to autonomous legal and ethical assessments to ensure that the concerns raised by these input parameters have been addressed (see Section 2.5 for more on this).

This indicates that technical and/or organisational measures to ensure that data sources presented for registration on the Platform can be lawfully (and in an ethical manner) leveraged by users have been implemented. As such, this Control is now fully implemented, as described in the final **WP4 Legal/Ethical Checklist** (see Annex 3 – Final WP4 Legal/Ethical Checklist).

2.3.4 Data Subject Rights

As reported in Section 2.3.4 of D3.6 [4], it is necessary to ensure that the Platform allows for the exercise of Data Subject Rights – at a minimum, the Platform should not create any relevant technical obstacles to the exercise of these rights.

It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the Platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered, regarding individuals whose personal data may be contained in data sources processed via the Platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the Platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of Data Subject Rights, concerning such individuals (for other relevant individuals, e.g., PolicyCLOUD users, see Section 2.2.4).

This indicates that technical and/or organisational measures to allow for the exercise of Data Subject Rights to individuals whose personal data may be processed on the Platform, concerning individuals whose personal data may be contained in relevant data sources, have been implemented. As such, this Control is now fully implemented, as described in the final **WP4 Legal/Ethical Checklist** (see Annex 3 – Final WP4 Legal/Ethical Checklist).

2.3.5 Data Quality (Minimisation, Retention, Accuracy)

As reported in Section 2.3.5 of D3.6 [4], the following steps had been taken concerning Controls no. 8 and 9 of the **WP4 Legal/Ethical Checklist**, as of the date of that Deliverable:

- As reported in Section 3.2.2.1 of D4.5 [14], mandatory and optional data constraints can be defined to configure the parameters under which the platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
- Various libraries are exploited in connection with the Platform's data validation, cleaning and verification activities to provide greater assurances of data quality (including accuracy, completeness and lack of errors).
- A mechanism for deletion of such personal data (regarding individuals whose personal data may be contained in data sources processed via the Platform – for PolicyCLOUD users, see Section 2.2.5) has been implemented within the Platform, and that the Platform does not present any relevant technical obstacles to the implementation of this ability and, consequently, to the possibility to enforce defined retention periods for such personal data.

As noted in Section 2.1, it has been agreed that the definition of retention periods for personal data contained within data sources processed via the Platform is the responsibility of the relevant UC Partner identifying each data source for processing (acting as controller for such personal data, under the GDPR). As a result, the issue of data retention is no longer considered an active Project-wide issue, and rather one which is to be dealt with at the individual UC level (see Section 2.5.2).

This indicates that technical and/or organisational measures to ensure that, where personal data is included in a data source to be processed via the Platform, only those personal data which are adequate, relevant and limited to what is necessary in relation to the specific purpose(s) for which the data source is to be processed are actually collected and further stored on the Platform (principle of data minimisation), that appropriate steps are taken to verify the accuracy of any such personal data and maintain it over time (principle of accuracy), and that such personal data are not retained beyond necessary for the mentioned purpose(s) (principle of storage limitation), have been implemented. As such, this Control is now fully implemented, as described in the final **WP4 Legal/Ethical Checklist** (see Annex 3 – Final WP4 Legal/Ethical Checklist).

2.4 Cross-sector Policy Lifecycle Management (WP5)

2.4.1 PolicyCLOUD user-facing tools (PDT and PME)

The PDT, as better described in D5.6 [15], is a web application consisting of a front-end that allows policy makers to evaluate PMs, and a back-end, that hides the complexity of storing the data models into persistent storage and implements the services that the front-end uses to display the content to the user and provide the necessary user experience. The PDT intentionally hides the complexity of the system dataflow to provide the user with a decision support system towards evidence based PPs. Concerning the back end, data storage and analytics have been addressed in Sections 2.2 and 2.3.

On the other end, the PME is a tool which links only with the Platform's front-end. The purpose of the PME is to allow the creation of new policies and editing of existing ones following an evaluation process. The PME helps the user to create PPs by applying relevant key performance indicators ("**KPIs**") that are stored into the PDT backend or compose new KPIs by defining actors and stakeholders to the PM, selecting the data source and the analytical tools that will process the data, and finally setting the parameter values that the analytical tools require to provide the result, and save the PM to the datastore. [15]

From a contractual perspective, as noted in D3.3 [3] and D3.6 [4], it is important to define T&Cs for the use of the PDT and PME, to properly regulate the service relationship established between the PolicyCLOUD manager(s) and tool users (i.e., individual users, or organizations to which the individual users belong). These T&Cs will need to be accepted for the use of the PDT and PME to be allowed. Matters regulated include:

1. Description of services offered through the PDT and the PME.
2. Payment terms.
3. Acceptable use of the PDT and the PME.
4. Intellectual property (in particular, ownership of the PDT and PME's assets, the input provided by users via the PDT and the PME, and the output generated by the PDT and the PME).
5. Data protection (with reference to the privacy policy[2]).
6. Warranties provided by the PolicyCLOUD manager(s) and liability terms related to provision of the PDT and the PME, considering the likelihood of statistical inaccuracies in PDT and PME output and the possibility for such output to be used to create public policies.

[2] See Section 2.4.2.

7. Warranties provided by PDT and PME users, regarding compliance with ethical and legal requirements around selection of data sources and use of the PDT and PME for policymaking purposes.
8. Applicable law.
9. Terms under which the T&Cs may be modified.
10. Termination and effects of termination.

The final status of implementation of the T&Cs concerning the PDT and the PME, as of the date of this Deliverable, is reported in the Updated WP5 Legal/Ethical Checklist (see Annex 4 – Updated WP5 Legal/Ethical Checklist).

2.4.2 Data Management

Information on PDT/PME users will typically be collected and stored in the PDT back-end for a variety of purposes (e.g., user authentication). This information is classifiable as personal data to the extent that it can be traced back to an individual user, and thus the following concerns have been addressed:

(1) *Lawfulness.* An appropriate legal basis for each purpose for which personal data on users may be collected has been identified, under Art. 6 GDPR. All steps needed to properly implement the identified legal bases have been taken, depending on the legal bases selected, which in turn depend on the way user personal data are processed via the PDT. The legal bases are:
   - Compliance with legal requirements, under Art. 6(1)(c) GDPR.
   - The legitimate interests of the PolicyCLOUD manager(s) and users (or the organizations to which users belong) to ensure the proper operation of the PDT/PME and overall Platform, under Art. 6(1)(f) GDPR.
   - The legitimate interests of the PolicyCLOUD manager(s) and users (or the organizations to which users belong) to ensure the security of the PDT/PME and overall Platform, as well as of data stored on the Platform, under Art. 6(1)(f) GDPR.
   - The legitimate interests of the PolicyCLOUD manager(s) in leveraging personal data as needed for the establishment, exercise, or defense of legal claims (e.g., brought against the PolicyCLOUD Manager(s) for damages resulting from use of the PDT/PME), under Art. 6(1)(f) GDPR.

(2) *Fairness.* Personal data on users is not used in a manner which may violate their reasonable expectations. Users are not profiled, and their data is not covertly shared with third parties for marketing purposes. Also, mechanisms are in place to ensure that users can exercise their rights in relation to any personal data of theirs which may be collected during the use of the PDT/PME. More specifically, each data subject can exercise the following rights by sending a request in writing to the PolicyCLOUD manager(s), to the extent permitted by applicable law:
   - To access. Data subjects can obtain information relating to the processing of their personal data and a copy of such personal data.
   - To erase. Data subjects can require the deletion of their personal data, to the extent permitted by law.

- To object. Data subjects can object to the processing of their personal data, on grounds relating to their situation. Where an objection is presented, the PolicyCLOUD manager(s) will be entitled to assess the request and refuse it if there are legitimate reasons to proceed with the processing that prevail over the freedoms, interests, and rights of the requesting data subject.
- To rectify. Where data subjects consider that their personal data is inaccurate or incomplete, they can require that such personal data be modified accordingly.
- To restrict. Data subjects can request the restriction of the processing of their personal data.

Furthermore, should data subjects believe that the processing of their personal data is contrary to the legislation in force, they have the right to lodge a complaint to the competent data protection supervisory authority.

(3) *Transparency.* A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR. Such privacy policy is attached to D8.1 [16]. More specifically, the data protection information notice is provided to PDT users using a two-layer approach:

- The first layer is a pop-up banner which is shown to users when accessing the PDT. This pop-up banner includes a link to the second layer data protection information notice (i.e., the extended version of the data protection information notice). The text of the pop-up banner is provided in D8.1 [16]. See also D3.6 [4] for more details.
- The second layer is the extended version of the data protection information notice, including all elements required by Arts 13 and 14 GDPR. This extended version of the data protection information notice, the text of which is provided under D8.1 [16], is accessible to PDT users by clicking on a footer named “End Users Data Protection Information Notice” published on all the pages of the PDT’s web environment. [4]

Also, the PDT and PME do not use cookies or similar tracking technologies, so no specific compliance requirements are needed to this regard.

(4) *Purpose limitation.* Personal data collected on PDT and PME users is used only for the purposes declared in the privacy policy, which are:

- Creation and management of the end user account and operation of the Platform.
- Ensuring the security and the proper use of the Platform.
- Defending PolicyCLOUD from legal and/or regulatory risks.

(5) *Data minimization.* Only the strict minimum amount of personal data needed for the above listed purposes are collected. More specifically, the personal data of PDT and PME users processed are:

- Name.
- E-mail address.
- Organization and role.
- Username and password.
- Internet protocol (“**IP**”) address.

- Event logs related to the use of the Platform.

(6) *Accuracy.* Users can rectify any personal data they submit, or which is collected on them, in connection with use of the PDT and PME, by writing to PolicyCLOUD.

(7) *Storage limitation.* Personal data on PDT and PME 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 should be irreversibly anonymized or deleted. This requires the identification of appropriate retention periods for any such personal data collected[3].

(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[4].

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

Furthermore, when offering the PDT and/or PME as a service, the PolicyCLOUD manager(s) 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 the PolicyCLOUD manager(s) and PDT users, as this would only cover the processor activities of the PolicyCLOUD manager(s). To comprehensively regulate this relationship, the arrangement will need to include obligations to govern the PolicyCLOUD manager(s) personal data processing activities as a controller. This can be done through a data management agreement, distinguishing between the different sets of processing activities performed and allocating 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.

Finally, the data visualization components of the PDT should 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. [3] [4] The core concerns around data visualization focus on:

(1) *Statistical accuracy.* Indeed, a failure to provide appropriate visualization results may lead to misleading or erroneous conclusions drawn by policymakers, culminating in misguided policy-making activities. As such, the Consortium ensures that enough testing is performed to ensure a reasonable degree of statistical accuracy for these activities. Any operations performed on data is methodically logged to ensure traceability, and the general rationale and logic behind the data visualization is explained to PDT users, so this can be considered during their decision-making process. Users are advised of the possibility of false positives or false negatives and errors in result presentation, so that they are incentivized to critically examine results produced by the

[3] See Section 2.2.5.

[4] See Section 2.1.1.

visualization functions in their decision-making process. The same considerations as developed for analytics functions in Section 2.3.1 apply, with the necessary adaptations.

(2) *Explainability of results.* For a PDT user to make use of the PDT in a meaningful way, as a tool to support policymaking while preserving their own accountability for the decisions taken regarding any specific policy, it is important that they can understand the output presented to them by the PDT. For this purpose, a User Manual has been drafted, giving clear explanations to users on:
    - Steps taken across the design and implementation of the PDT and overall Platform to ensure compliance with applicable legal and ethical requirements, those implemented to maximize the security and robustness of the Platform, as well as the accuracy and reliability of 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 user.

The status of implementation of data management requirements concerning the PDT/PME, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the Updated WP5 Legal/Ethical Checklist (see Annex 4 – Updated WP5 Legal/Ethical Checklist).

2.5 Use Cases Adaptation, Integration & Experimentation (WP6)

2.5.1 Data Source Selection

2.5.1.1 UC #1: Participatory Policies against Radicalization

UC #1 develops a collaborative data-driven analysis for the validation of existing policies to counter radicalization based on a participatory review of data coming from social media (e.g., Twitter, Reddit, Really Simple Syndication feeds) and open datasets (e.g., GTD and RAND Database of Worldwide Terrorism Incidents – "**RDWTI**"). In addition, it provides useful insights and valuable information to policy makers at any level (local, regional, and national) to update current policies and/or create new ones, while at the same time allow them to interact with other relevant stakeholders (i.e., law enforcement agencies – "**LEAs**" -, social services, schools, 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 [17].

UC #1's objective is to improve operational efficiency, transparency, and decision making by using the Platform. The PolicyCLOUD visualization technologies will enable policy makers to identify issues, trends, and policy effects and interactions, while the analytics technologies will enable them to discover insights and find meaningful explanations about the effects of policies [17].

Scenario A: Radicalization Incidents

UC #1's Scenario A aims to monitor the occurrence of radicalization incidents in the geographic proximity of a region. Data coming from the GTD and RDWTI are used. The policy maker can select an area of interest and consult the different incidents that have taken place in each period. The goals of this scenario are to validate existing policies and investigate if there is a need to adjust or update them or create a new one based on the retrieved information [17].

For this scenario, the relevant data sources [18] are:

- Dataset 7 (GTD), an open-source database including information on domestic and international terrorist attacks around the world taking place between 1970 and 2018, which now includes more than 190,000 cases [19].
- Dataset 8 (RDWTI), an open-source database including information of international and domestic terrorism incidents from 1968 through 2009, which now includes more than 40,000 cases [20].

The use of these data sources does not raise specific legal, regulatory, ethical, or societal concerns, provided that the data sources are used in accordance with their respective licensing T&Cs:

- The GTD's End User License Agreement ("EULA") defines the conditions under which the GTD can be accessed and used to conduct research and analysis for non-commercial purposes [21]. It is necessary to ensure that the use of this data source, in the context of PolicyCLOUD, is strictly aligned with the conditions set by this EULA.
- The RDWTI is publicly available. As stated on RAND's website, everyone is welcome to use this database; however, all public use of these data must be attributed to the RDWTI [22]. It is necessary to ensure compliance with this attribution requirement.

The implementation status of requirements concerning UC #1's Scenario A, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 5 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario A (Radicalization Incidents)).

Scenario B: Radicalized Groups and Individuals

UC #1's Scenario B aims to identify the main actors (individuals or groups) involved in extremism activities or propaganda spreading, through online and offline activities. Data coming from the GTD and RDWTI are used. The Policy Maker can select a set of keywords to identify the individuals and groups active in the area of their interest and consult the different incidents linked to each of them. As with Scenario A, the goals of this scenario are to validate existing policies and investigate if there is a need to adjust or update them or create a new one based on the retrieved information [17].

For this scenario, the relevant data sources are the same as in Scenario A [18]. The same considerations as to applicable legal, regulatory, ethical, and societal requirements as drawn for Scenario A apply.

The implementation status of requirements concerning UC #1's Scenario B, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 6 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario B (Radicalized Groups and Individuals)).

SCENARIO C: TREND ANALYSIS

UC #1's Scenario C aims to identify current and future trends related to radicalization efforts, through keyword detection, new entity recognition and new term identification. Data collected from social media will be used. The policy maker can select keywords of interest and consult the different information linked to them. As with Scenarios A and B, the goals of this scenario are to validate existing policies and investigate if there is a need to adjust or update them or create a new one based on the retrieved information [17].

For this scenario, the relevant data source is Twitter (i.e., Dataset 4: relevant Twitter posts published by users, captured, and processed for subsequent analysis in UC #1) [18].

From a legal, regulatory, ethical, and societal standpoint, the use of this data source, in the context of Scenario C, raises the following points of attention:

(1) Twitter's terms of service [23] specify that a third party may not access or search or attempt to access or search Twitter's services by any means (automated or otherwise) other than through currently available, published interfaces provided by Twitter (and only pursuant to the applicable T&Cs), unless the third party has been specifically allowed to do so in a separate agreement with Twitter. Also, crawling Twitter's services is permissible if done in accordance with the provisions of Twitter's robots.txt file[5]; however, scraping Twitter's services without Twitter's prior consent is expressly prohibited.

(2) The UC Partner should assess whether this data source, or any relevant parts of this data source which are to be processed via the PolicyCLOUD platform under this scenario, may be qualified as a protected database under Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (the "**Database Directive**"), under copyright, or under sui generis protection, under the local laws applicable to the UC Partner (or the PolicyCLOUD users for which this scenario is intended). Where this is the case, the UC partner should assess whether any authorizations from relevant rightsholders (e.g., Twitter, Twitter users) is required under those laws, and obtain such authorizations if needed.

(3) The information available related to this scenario strongly implies that the leveraging of this data source will involve the processing of personal data. This will be the case where tweets, or other Twitter content, are collected in a form allowing their poster, uploader, or sharer to be identified (e.g., through their name, online handle, or other potential identifiers included in the content).

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

The relevant privacy and data protection compliance obligations which may be triggered as a result are further described in Section 2.5.2.

(4) The UC Partner should take measures to assure itself of the reliability of this data source. In particular, the likelihood that any data collected from this data source may be false, inadequate, inaccurate, or incomplete (also in terms of representativeness of the target population), considering the goals of this scenario, must be carefully assessed. Any relevant reliability and accuracy concerns detected must be identified, and measures taken to address those concerns must be appropriately documented.

The implementation status of requirements concerning UC #1's Scenario C, as of the date of this Deliverable, and next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 7 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario C (Trend Analysis)).

Scenario D: (Near) Real-time Assessment of Online Propaganda

UC #1's Scenario D aims at providing the ability to understand the context of specific events and online activities (sentiment analysis, opinion mining, location surveillance, and user monitoring). Data collected from social media are used. As with Scenario C, PolicyCLOUD users can select keywords of interest and consult information linked to those keywords. As with the remaining scenarios of UC #1, the policy maker can select keywords of interest and consult the different information linked to them. As with the remaining scenarios of UC #1, the goals of this scenario are to validate existing policies and investigate if there is a need to adjust or update them or create a new one based on the retrieved information [17].

For this scenario, the relevant data source is the same as in Scenario C [18]. The same considerations as to applicable legal, regulatory, ethical, and societal requirements as drawn for Scenario C apply.

The implementation status of requirements concerning UC #1's Scenario D, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 8 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario D ([Near] Real-time Assessment of Online Propaganda)).

2.5.1.2 UC #2: Intelligent Policies for the Development of Agri-food Industry

UC #2 aims at achieving the following main objectives [17]:

(1) Improving investments in agri-food promotion by the Government of Aragon.
(2) Facilitating tools and access to innovative technologies for small and medium producers, tools based on open data, social media analysis, and opinion mining.
(3) Improving the distribution of products thanks to the tools created through PolicyCLOUD, which will allow the searching and comparing of prices, as well as the positioning of PolicyCLOUD users' products and those of their competitors.
(4) Supporting decision-making regarding investment in different geographical areas with market study elements based on AI.

(5) Bringing the agri-food industry closer to innovative technologies.
(6) Supporting policy makers in the design and modelling of new policies and/or updating existing ones.
(7) Creating stable working groups between producers and policy makers allowing the improvement of communication channels and the development of new tools.

Scenario A.1: Politika Price Point

UC #2's Scenario A.1 aims to define the base case for wines X and Y as a set of parameters and their values. The base case reflects the current data for the two wines examined. Furthermore, the policy maker defines different alternatives in terms of a set of parameters and their values that are different from the ones in the base case and look promising in their potential for increasing the popularity of X. The goal of this scenario is to allow the policy maker to provide data for two competing wines, one from Aragon (e.g., X) and one from some other region (e.g., Y) that describe their current price, quality, and advertisement effort. Policy maker can then support the wineries to define pricing alternatives and advertisement effort for Aragon wines to make those more popular among a specific population. Simulation results are visualized as charts facilitating direct comparisons between the alternatives explored [17].

From a legal, regulatory, ethical, and societal standpoint, the data sources used in this scenario raise the following points of attention:

(1) The relevant UC Partner should carefully assess T&Cs applicable to each website selected as a data source, to ensure that data collected from those data sources can be leveraged for the purposes of this scenario. A thorough analysis of the general and specific T&Cs linked to each data source is necessary, to ensure that those data sources can be leveraged for these purposes without breaching any contractual obligations.
(2) Similarly, the UC Partner should assess whether these data sources, or any relevant parts of these data sources which are to be processed via the Platform under this scenario, may be qualified as a protected database under the Database Directive, under copyright, or under sui generis protection, under the local laws applicable to the UC Partner (or the PolicyCLOUD users for which this scenario is intended). Where this is the case, the UC partner should assess whether any authorizations from relevant rightsholders (e.g., owners of the websites) is required under those laws and obtain such authorizations if needed. This assessment will need to be conducted per each individual data source.
(3) The leveraging of these data sources (or of some of these data sources) involves the processing of personal data. The relevant privacy and data protection compliance obligations which may be triggered as a result are further described in Section 2.5.2.
(4) The UC Partner should take measures to assure itself of the reliability of this data source. In particular, the likelihood that any data collected from this data source may be false, inadequate, inaccurate, or incomplete (also in terms of representativeness of the target population), considering the goals of this scenario, must be carefully assessed. Any relevant reliability and

accuracy concerns detected must be identified, and measures taken to address those concerns must be appropriately documented.

The implementation status of requirements concerning UC #2's Scenario A.1, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 9 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario A.1 (Politika Price Point)).

Scenario A.2: Price Evolution

UC #2's Scenario A.2 aims to visualize the sale price of wine on different specialized websites, with automatic warning systems that avoid penalties for contracts with large distributors. The goal of this scenario is to support the improvement of commercial policy for the regional agri-food industry [17].

For this scenario, the same considerations as to applicable legal, regulatory, ethical, and societal requirements as drawn for Scenario A.1 apply.

The implementation status of requirements concerning UC #2's Scenario A.2, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 10 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario A.2 (Price Evolution)).

Scenario B: Opinion on Social Networks

UC #2's Scenario B aims at visualizing the negative and positive opinions on social networks of the various products analyzed allowing an automatic and immediate response to the end user. The goal of this scenario is to create an immediate communication with the end user, knowing their impressions, both positive and negative, that will allow to interact with the end customer more directly [17].

From an ethical, legal, regulatory, and societal standpoint, the data sources used in this scenario raise the following points of attention:

(1) The relevant UC Partner should carefully assess T&Cs applicable to the social media platform ("**SMP**"), to ensure that data collected from this data source can be leveraged for the purposes of this scenario. A thorough analysis of the general and specific T&Cs linked to the data source (i.e., general terms of use, or more specific terms of use which may be available for use of the relevant platform for professional purposes, potentially covering the use of APIs connected to the platform) is necessary, to ensure that the data source can be leveraged for these purposes without breaching any contractual obligations.

(2) In particular, Twitter's terms of service [23] specify that a third party may not access or search or attempt to access or search Twitter's services by any means (automated or otherwise) other than through currently available, published interfaces provided by Twitter (and only pursuant to the applicable T&Cs), unless the third party has been specifically allowed to do so in a separate agreement with Twitter. Also, crawling Twitter's services is permissible if done in accordance with

the provisions of Twitter's robots.txt file[6]; however, scraping Twitter's services without Twitter's prior consent is expressly prohibited.

(3) The UC Partner should assess whether these data sources, or any relevant parts of these data sources which are to be processed via the Platform under this scenario, may be qualified as a protected database under the Database Directive, under copyright, or under sui generis protection, under the local laws applicable to the UC Partner (or the PolicyCLOUD users for which this scenario is intended). Where this is the case, the UC partner should assess whether any authorizations from relevant rightsholders (e.g., SMP providers or users) is required under those laws and obtain such authorizations if needed.

(4) The leveraging of the data source related to this scenario involves personal data processing. This is the case where posts, or other social media content, are collected in a form which allows their poster, uploader, or sharer to be identified (e.g., through their name, online handle, or other potential identifiers included in the content). The relevant privacy and data protection compliance obligations which may be triggered as a result are further described in Section 2.5.2.

(5) The UC Partner should take measures to assure itself of the reliability of this data source. In particular, the likelihood that any data collected from this data source may be false, inadequate, inaccurate, or incomplete (also in terms of representativeness of the target population), considering the goals of this scenario, must be carefully assessed. Any relevant reliability and accuracy concerns detected must be identified, and measures taken to address those concerns must be appropriately documented.

The implementation status of requirements concerning UC #2's Scenario B, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 11 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario B (Opinion on Social Networks)).

SCENARIO C: TREND ANALYSIS

UC #2's Scenario C aims to analyze the trends in the wine sector through the specialized websites of the sector. The goal of this scenario is allowing to know the trends in each of the markets of interest; knowing the trends in the sector allows to adjust the diffusion policies considering all parameters [17].

From an ethical, legal, regulatory, and societal standpoint, the data sources used in this scenario raise the same points of attention as described for UC #2's Scenarios A.1 and A.2, with the necessary adaptations.

The implementation status of requirements concerning UC #2's Scenario C, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the

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

relevant Updated WP6 Legal/Ethical Checklist (see Annex 12 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario C (Trend Analysis)).

2.5.1.3 UC #3: FACILITATING URBAN POLICY MAKING AND MONITORING THROUGH CROWDSOURCING DATA ANALYSIS

UC #3 aims at achieving the following main objectives [17]:

(1) Supporting the UC Partner's urban policy making in key areas for the city of Sofia.
(2) Supporting PolicyCLOUD users in the design and modelling of new policies and/or updating of existing policies, by enhancing their knowledge and capability to identify problems, issues and trends related to urban environments.
(3) Supporting policy makers in the adoption or modification of adequate policy making decisions on budget planning and effective use of budget and public resources.
(4) Enhancing the UC Partner's operational efficiency, transparency, and decision-making, to improve the overall quality of life in the city of Sofia.
(5) Using visualization tools to support PolicyCLOUD users in identifying issues, trends, and tendencies (including behavioral trends), and policy effects and interactions.
(6) Validating existing policies and assess potential policies and initiatives in focus areas.
(7) Facilitating more efficient use of resources.
(8) Sharing best practices and lessons learned with relevant stakeholders (at any level).
(9) Identifying tendencies.
(10) Facilitating better control, monitoring, and prevention.
(11) Creating new policies based on the retrieved information.

SCENARIO A: VISUALISATION

UC #3's Scenario A aims to:

- visualize the signals received via Sofia's Call Centre CallSofia related to:
  (1) road infrastructure.
  (2) environment and air quality.
  (3) waste collection and waste disposal.
  (4) transport and parking.
  (5) cleanliness of public spaces.
  (6) violation of public order.
- provide a detailed analysis of their frequency over time and territorial distribution by categories, types, areas, districts, etc.
- support and facilitate data-based municipal decision-making [17].

For this scenario, the relevant data source is Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality [18].

From a legal, regulatory, ethical, and societal standpoint, the data source used in this scenario raises the following points of attention:

(1) The leveraging of this data source involves the processing of personal data. The relevant privacy and data protection compliance obligations which may be triggered as a result are further described in Section 2.5.2.

(2) The UC Partner should take measures to assure itself of the reliability of this data source. In particular, the likelihood that any data collected from this data source may be false, inadequate, inaccurate, or incomplete (also in terms of representativeness of the target population), considering the goals of this scenario, must be carefully assessed. Any relevant reliability and accuracy concerns detected must be identified, and measures taken to address those concerns must be appropriately documented.

Given that the UC Partner is also responsible for the management of the data source, there are no applicable contractual or intellectual property-related restrictions applicable to the leveraging of this data source in this scenario.

The implementation status of requirements concerning UC #2's Scenario A, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 13 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario A (Visualization)).

Scenario B: Predictive Analysis

UC #3's Scenario B aims to use predictive analysis to predict a future outcome. The goal of this scenario is to use specially designed algorithms from PolicyCLOUD to predict future outcomes and trends with regards to the key sectors already mentioned with regards to Scenario A (potential challenges, effects of new policies, etc.). using the provided datasets [17].

For this scenario, the relevant data source is Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality [18]. From a legal, regulatory, ethical, and societal standpoint, the identified data source raises the same points of attention as in Scenario A.

The implementation status of requirements concerning UC #2's Scenario B, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 14 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario B (Predictive Analysis)).

Scenario C: Environment and Air Quality Cross Analysis

UC #3's Scenario C aims to provide a cross-analysis of the datasets on environment and air quality received via Sofia's call center Call Sofia and the additional data provided by Sofia's air quality measurement stations to support and facilitate data-based municipal decisions in one of the key areas of urban development [17].

For this scenario, the relevant data sources are [18]:

- Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality.
- Dataset 13 (Sofia Municipality Air Quality Sensors Data), i.e., real time Internet of Things sensors data for monitoring and measurement of the air quality.

From a legal, regulatory, ethical, and societal standpoint, the first data source identified (Dataset 12) raises the same points of attention as in Scenarios A and B. As for the second data source identified (Dataset 13), the UC Partner should assure itself of the reliability of this data source. In particular, the likelihood that any data collected from this data source may be false, inadequate, inaccurate, or incomplete (also in terms of representativeness of the target population), considering the goals of this scenario, must be carefully assessed. Any relevant reliability and accuracy concerns detected must be identified, and measures taken to address those concerns must be appropriately documented.

Concerning Dataset 13, from a contractual standpoint, the UC Partner is authorized to use the data in the context of this scenario. Additionally, leveraging this data source does not imply personal data processing.

The implementation status of requirements concerning UC #2's Scenario C, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklist (see Annex 15 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario C (Environment and Air Quality Cross Analysis)).

2.5.2 Privacy and Data Protection Compliance

2.5.2.1 UC #1: Participatory policies against Radicalization

In the context of UC #1, Scenarios C and D raise potential privacy and data protection concerns, since the relevant data source identified for these scenarios (i.e., Dataset 4: relevant Twitter posts published by users, captured, and processed for subsequent analysis in UC #1) strongly implies that the processing of personal data will be involved [18]. For example, the collection and further processing of the following categories of personal data may be relevant to these Scenarios:

(1) Names, surnames and/or pseudonyms (e.g., online handles).
(2) Information publicly disclosed on Twitter (which may contain personal data).
(3) Profiles created on Twitter users, based on information inferred from the automatic processing of the above categories of personal data.

In this Section, we will present an overview of the main issues which the UC Partner (or intended PolicyCLOUD user) – acting as a controller[7], as it will leverage the Platform (and associated personal data processing activities) for its own specific purposes – must bear in mind when defining PolicyCLOUD platform use requirements in the context of these scenarios, with reference to the GDPR's principles. [3] [4]

The implementation status of privacy and data protection requirements concerning UC #1, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklists (see Annexes 5, 6, 7 and 8). Furthermore, dedicated DPIAs concerning UC #1 Scenarios C and D are provided under Annexes 16 and 17.

LAWFULNESS

Any use of personal data must be performed based on some 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 [3] [4].

Indeed, although personal data processed for these scenarios may be publicly available on Twitter, this does not exempt the UC Partner (or the intended PolicyCLOUD users) from the obligation to find a suitable legal basis for the processing of such data for its (their) own specific purposes. An assessment as to which of the legal bases afforded by the GDPR may be applicable and implementable in the context of these scenarios must therefore be conducted. This assessment must consider the full context of the scenarios, including the specific data sources to be used (e.g., the types of personal data included in those data sources, the way 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 the scenarios.

To this end, considering the characteristics of Dataset 4 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on the particularities of specific legal bases which may potentially be relevant for consideration.

2.5.2.1.1.1 CONSENT

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 PolicyCLOUD users to interact directly with each specific data subject which may upload content to Twitter which is considered relevant by the Platform's algorithms before the processing of such content is carried out, to request

[7] After completion of PolicyCLOUD, the PolicyCLOUD manager(s), on the other hand, will arguably act as a processor on behalf of the abovementioned controller, as the activities under the PolicyCLOUD manager(s)'s responsibility will be limited to making personal data available for processing by the controller via the PolicyCLOUD platform, in strict compliance with the controller's instructions and specifications, for the specific purposes of the controller.

their consent for the processing of any of their personal data which may be collected. As such, other alternatives should be explored.

2.5.2.1.1.2 Legitimate interests

Art. 6(1)(f) GDPR is not a legal basis which the GDPR affords to the personal data processing carried out by public authorities in the performance of their tasks [3] [4]. As such, if a PolicyCLOUD 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 that user. To the extent that this is the case, this legal basis is not available in the context of this UC.

2.5.2.1.1.3 Public interest

Where it is not feasible for a PolicyCLOUD user to rely on either of the above options, that 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 [3] [4].[8] As this legal basis presents the most flexible approach available to public authorities, when acting in the performance of their tasks, each user should assess whether the requirements described in D3.3 [3] and D3.6 [4] for this legal basis are met, to ensure the lawfulness of the intended processing activities.

Lawfulness (Special Categories of Personal Data)

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[9], from those listed in Art. 9(2) GDPR, or as may be further provided under applicable Member State law, must also be identified [3] [4]. 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 characteristics of Dataset 4 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on the particularities of specific derogations which may potentially be relevant for consideration.

[8] Art. 6(1)(e) GDPR.

[9] Established by Art. 9(1) GDPR.

2.5.2.1.1.4 Explicit consent

On this point, we refer to the requirements set out in D3.3 [3] and D3.6 [4], as those must be met to ensure validity of consent, and on the general observations on the viability of reliance on consent for this use case made in Section 2.5.2.1.1.1.

2.5.2.1.1.5 Data manifestly made public by the data subject

As noted in D3.3 [3] and D3.6 [4], 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.[10] This is a particularly relevant derogation for this UC, to the extent that personal data is to be collected from publicly available content uploaded on SMPs. However, in this regard, the factors indicated by the EDPB, in the context of SMPs, for determining whether this derogation is met are particularly relevant. The following points should be borne in mind:

(1) If a SMP's default setting, for the publication of content, are private, such that a data subject needs to actively choose to share data publicly, this weighs in favor 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 in this UC.

(2) Where a SMP is meant for the public disclosure of data in a non-intimate or personal setting, such as is the case with Twitter, this weighs in favor of this derogation.

(3) If SMPs inform their users that their content may be collected by other organizations, for purposes identical or similar to those pursued by the end-user, this weighs in favor of this derogation. Twitter's Privacy Policy [24] 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 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*". Each PolicyCLOUD user should make its own privacy policy publicly-available and easily accessible on its website(s) to further increase the likelihood of social media users becoming aware of how their data may be used [3] [4].

Each PolicyCLOUD user should carefully assess whether this derogation may be applicable to any special categories of personal data (if any are collected) extracted from Twitter.

2.5.2.1.1.6 Substantial public interest

[10] Art. 9(2)(e) GDPR.

This derogation requires a demonstrable need to process special categories of personal data to meet a substantial public interest with a basis in EU or Member State law applicable to the controller (i.e., the PolicyCLOUD user), which must [3] [4]:

- 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.

Each PolicyCLOUD user should assess whether this derogation may be applicable, considering the purposes and goals relevant to this UC.

2.5.2.1.1.7 Statistical purposes

In light of the requirements explained in D3.3 [3] and D3.6 [4], as the processing of special categories of personal data, if relevant, may be carried out in this UC for statistical purposes, each PolicyCLOUD user should determine whether this processing can be based on EU or Member State law applicable to that user, 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.

Where this is the case, this derogation may be applicable, provided that the processing carried out by the relevant PolicyCLOUD user via the Platform is supported by specific safeguards, as specified in Art. 89(1) GDPR, and further described in D3.3 [3] and D3.6 [4]. To provide assurances that this derogation may be applicable to the relevant PolicyCLOUD users, the UC Partner and the Consortium must collaborate to ensure that such safeguards are implemented on the Platform.

Fairness

To consider the personal data processing as fair under the GDPR, the controller must ensure that personal data are handled in ways that may be reasonably expected by data subjects, and must not use such data in a way that may produce unjustified adverse effects on them [3] [4].

Considering Dataset 4 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on specific key elements to be considered by the UC Partner, under the fairness principle.

2.5.2.1.1.8 INTERACTION

Since PolicyCLOUD collects data indirectly from Twitter, this requires the relevant PolicyCLOUD 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 D3.3 [3] and D3.6 [4], this also requires the 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 UC, such as by publishing information on these activities on a publicly available website managed by the user.

2.5.2.1.1.9 EXPECTATION

This element is particularly important in this UC, as depending on what personal data about the data subjects is made available to the relevant PolicyCLOUD user for processing via the Platform, and the subsequent processing actions which the user may perform on those data, the fairness of such processing may be called into question (due to being reasonably unexpected by the relevant data subjects) pursuant to the GDPR. In particular, while it may be argued that the collection of data subjects' publicly available comments, posts, and interactions on Twitter and their subsequent aggregation to provide relevant statistics, trends, charts, etc., may, to some extent, be expected by data subjects when they publicly post or comment on Twitter (or, at least, could arguably be seen as not excessively intrusive), the same is arguably not the case if the Platform were to, e.g., allow for the storage of data subjects' opinions (posted on social media) in such a way that allows the PolicyCLOUD user to directly identify data subjects, trace content back to its original source (e.g., through a hyperlink) and directly interact with those data subjects. This level of intrusiveness may potentially be considered as excessive and, as such, run afoul of this principle. Furthermore, such activity could be considered as out-of-scope in the context of the policy-making purposes for which the PolicyCLOUD platform is being developed. As such, each relevant PolicyCLOUD user should consider, whenever possible, the aggregation of personal data collected from public sources, to mitigate the risk of unfair processing.

One exception 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) suggests that they may have a lessened expectation of privacy regarding content which they make publicly available on social networks, such that the processing of such content by PolicyCLOUD users for the purposes pursued in this UC would not be an unreasonable violation of their expectations. This should be carefully assessed by each relevant PolicyCLOUD user when identifying the legal basis (or bases) applicable to them (e.g., in the LIA 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).

2.5.2.1.1.10 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.

TRANSPARENCY

To ensure compliance with the transparency principle, each relevant PolicyCLOUD user must ensure that it provides complete and understandable information to the relevant data subjects on their data processing practices. [3] [4]

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 UC (but indirectly, from Twitter), PolicyCLOUD users may be able to argue for the exemption provided under Art. 14(5)(b) GDPR, where the provision of information directly to each individual social media user proves impossible or would represent a disproportionate effort for those PolicyCLOUD users [25]. Each PolicyCLOUD user should develop a specific assessment to demonstrate that this exception is applicable, contrasting the effort required of the PolicyCLOUD 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 PolicyCLOUD user). This can be conducted as part of a broader DPIA. Where the effort required of a PolicyCLOUD user would be clearly disproportionate considering the (low) impact upon data subjects, this exemption can be relied on. Where this exemption applies, PolicyCLOUD users must act appropriately to ensure the protection of the rights and freedoms of social media users although this information is not directly provided to them, such as by displaying the information on a publicly available website, as stated in Art. 14(5)(b) GDPR. Any information notice or privacy policy developed (e.g., to be made available on a website managed by a PolicyCLOUD 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 prioritized, what the appropriate level of detail is and which are the best means by which to convey this information to data subjects [26]. Whenever feasible, PolicyCLOUD users 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 [26] [27].

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

PURPOSE LIMITATION

To ensure compliance with the principle of purpose limitation, under Art. 5(1)(b) GDPR, PolicyCLOUD users must identify specific, explicit, and legitimate purposes for which personal data are to be collected and processed via the Platform, and then refrain from using personal data for any other incompatible purpose [3] [4]. 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(1)(b) GDPR, provided that the safeguards of Art. 89(1) GDPR are respected, it is arguably possible for users to make further use of personal data collected in a commercial context (such as in the context of SMPs) for statistical analysis aimed at the pursuit of a public interest.

To this end, considering the characteristics of Dataset 4 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on specific key elements to be considered by the UC Partner, under the purpose limitation principle.

2.5.2.1.1.11 SPECIFICITY

The specific purposes for which the PolicyCLOUD user intends to process data collected from SMPs should be made clear in the user's publicly available information notice, as mentioned in D3.3 [3] and D3.6 [4].

2.5.2.1.1.12 NECESSITY

The PolicyCLOUD user should carefully assess whether the purposes for which it intends to use the Platform can be met by using 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).

2.5.2.1.1.13 COMPATIBILITY

The presumption of compatibility mentioned above arguably applies to this UC, for research and statistical purposes.

2.5.2.1.1.14 LIMITATIONS TO FURTHER PROCESSING

The point raised in Section 2.5.2.1.1.9 on tracking or targeting of specific social media users is relevant for this element also.

DATA MINIMISATION

Compliance with the data minimization principle requires a minimalistic approach to personal data [3] [4], in the sense that:

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

- Only personal data which are adequate, relevant, and strictly necessary to meet a purpose should be used.
- 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 Dataset 4 [18], aside from the general guidance provided in D3.3 [3] and D3.6 [4], the UC Partner should arguably assess, for example, whether there is any added value in retaining the ability to identify specific users uploading content onto Twitter for the policy-making purposes which are pursued by this UC. If so, the UC Partner 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 the UC Partner determines that the use of personal data, preserving a link to the identity of individual social media users, is necessary, then the UC Partner 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.

ACCURACY

Ensuring accuracy of data used is fundamental from the legal perspective, but also from the ethical perspective, since 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 [3] [4].

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

To this end, considering the characteristics of Dataset 4 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on specific key elements to be considered by the UC Partner, under the principle of accuracy.

2.5.2.1.1.15 DATA SOURCE RELIABILITY

Where SMPs are concerned, reliability of content uploaded will always be a relevant risk to be mitigated (as those platforms do not exercise any editorial powers over the content uploaded by social media users). The UC Partner, in collaboration with the Consortium, 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.

2.5.2.1.1.16 Verification

Since this UC relies on several streaming data sources, a continuous effort to ensure accuracy must be conducted, namely by overwriting older data with newer data, to prevent excessive data aging, through specific and adequate "rollover" periods to be determined by the UC Partner.

2.5.2.1.1.17 Updated data

Whenever a streaming data source is used – which is the case for Twitter – appropriate rollover periods (i.e., periods after which older data must be overwritten by newer data collected) should be defined.

Storage limitation

This principle becomes relevant whenever personal data is collected, if it is not promptly anonymized or aggregated (with the underlying raw data being deleted as soon as possible) [3] [4]. Noting that personal data avoidance is the preferred approach under the data minimization principle, whenever this is not feasible, then the relevant PolicyCLOUD user should define specific retention periods for the personal data collected, based on the strict minimum period for which retention of those data is needed to ensure that the purpose for their collection and processing can be met.

Considering Dataset 4 [18], aside from the general guidance provided in D3.3 [3] and D3.6 [4], the UC Partner should consider that, when dealing with streaming data sources, it is important for the PolicyCLOUD users to be able 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 its users to define retention periods and include tools for automatic deletion or aggregation of underlying data after a user-defined retention period is exceeded.

2.5.2.2 UC #2: Intelligent policies for the development of agri-food industry

In this Section, we will present an overview of the key issues which the UC Partner (or intended PolicyCLOUD user) – acting as a controller – must bear in mind when defining PolicyCLOUD platform use requirements in the context of UC #2, with reference to the GDPR's principles described in D3.3 [3] and D3.6 [4].

Of the different data sources indicated for this UC [18], Dataset 9 (Wine varieties and brands information from Twitter) presents a clear potential for capturing personal data. As such, this Section will, in principle, be more relevant to this specific data source.

It should first be noted that this specific UC [17] may entail the collection and further processing of various categories of personal data, such as:

- Names, surnames and/or pseudonyms (e.g., online handles).

- Individual opinions posted on Twitter and – more generally – information publicly disclosed on Twitter (which may contain personal data).
- Profiles created on Twitter users, based on information inferred from the automatic processing of the above categories of personal data.

For the purposes of this UC, the UC Partner or intended PolicyCLOUD 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. Given that the dataset relevant to this UC which may contain personal data is comparable to the dataset relevant to UC #1 which may contain personal data (as all such datasets refer to information extracted from Twitter), the considerations drawn in Section 2.5.2.1 may be extended to UC #2, with the necessary adaptations and with the below specifications.

The status of implementation of privacy and data protection requirements concerning UC #2, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklists (see Annexes 9, 10, 11 and 12). Furthermore, dedicated DPIAs concerning UC #2 are provided under Annexes 16, 17,18 and 19.

Lawfulness

2.5.2.2.1.1 Legitimate interests

Since the UC Partner or intended PolicyCLOUD users, in leveraging the PolicyCLOUD platform, may be seen as operating as if they were a private entity, the legal basis afforded by Art. 6(1)(f) GDPR is available.

Therefore, each PolicyCLOUD user will need to perform an LIA, following the practical steps described in D3.3 [3] and D3.6 [4]. This has been conducted as part of a broader DPIA, which is provided Annexes 16, 17 18 and 19. Where the interests of the 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.

Purpose Limitation

2.5.2.2.1.2 Compatibility

The presumption of compatibility mentioned above applies here, for research and statistical purposes. However, considering the general scope of the Platform and the purposes and goals of this UC, the UC Partner or intended PolicyCLOUD users should generally refrain from further processing personal data to track or target specific social media users and directly interact with them on SMPs, as this purpose may arguably be considered as excessive and unlawful (thereby incompatible with the other purposes for which the UC Partner, or intended PolicyCLOUD users, may wish to use those data), as specified under Section 2.5.2.1.1.9.

2.5.2.3 UC #3: Facilitating urban policy making and monitoring through crowdsourcing data analysis

Of the two data sources relevant for UC #3 [18], the only one to present a clear potential for capturing personal data is Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality. In this Section, we will present an overview of the main issues which the UC Partner – acting as a controller – must bear in mind when defining the Platform use requirements in the context of the use of Dataset 12, with reference to the GDPR's principles described in D3.3 [3] and D3.6 [4].

The status of implementation of privacy and data protection requirements concerning UC #3, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the relevant Updated WP6 Legal/Ethical Checklists (see Annexes 13, 14 and 15). Furthermore, dedicated DPIAs concerning UC #3 are provided under Annexes 20, 21 and 22.

Lawfulness

The processing of special categories of personal data is carried out in UC #3 for statistical purposes and according with Art. 25m of the Bulgarian Personal Data Protection Act, which states that "*Personal data originally collected for a different purpose may be processed for the purposes of the National Archive Funds, for scientific, for historical research or for statistical purposes. In such cases, the controller shall apply appropriate technical and organizational measures to safeguard the rights and freedoms of the data subject in accordance with Article 89 (1) of Regulation (EU) 2016/679.*" [3] [4].

Fairness

For the processing under this UC to be considered as fair under the GDPR, the UC Partner must ensure that personal data are managed in ways that may be expected by data subjects and not use such data in a way that may produce unjustified adverse effects on them [3] [4].

To this end, considering the characteristics of Dataset 12 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on specific key elements to be considered by the UC Partner, under the fairness principle.

2.5.2.3.1.1 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.

2.5.2.3.1.2 Non-discrimination

The UC Partner must not discriminate against data subjects. 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 [3] [4].

2.5.2.3.1.3 Power balance

Asymmetric power balances shall be avoided or mitigated when possible. The UC Partner 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.

Transparency

To ensure its compliance with the principle of transparency, the UC Partner must ensure that it provides complete and understandable information to data subjects on its data processing practices [3] [4].

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 UC Partner 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 prioritized, what the appropriate level of detail is, and which are the best means by which to convey this information to data subjects [25]. Whenever feasible, the UC Partner 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. [25] [26] 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 center, including a link to the more complete privacy statement made available online [25].

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 [25]. 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 [25].

Therefore, with regards to the data protection information notice published on the Sofia Municipality website [28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the* *development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task carried out in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR.*".

Purpose Limitation

To ensure compliance with the principle of purpose limitation, under Art. 5(1)(b) GDPR, the UC Partner 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 [3] [4]. 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(1)(b) GDPR, provided that the safeguards of Art. 89(1) GDPR are respected, it is possible for the UC Partner to make further use of personal data collected in the context of the municipality contact center to be further used for statistical analysis aimed at the pursuit of a public interest [27].

To this end, considering the characteristics of Dataset 12 [18], further guidance to complement that of D3.3 [3] and D3.6 [4] can be provided on specific key elements to be considered by the UC Partner, under the principle of purpose limitation.

2.5.2.3.1.4 Specificity

The purposes must be specific to the processing; it should be made explicitly clear to citizens why their personal data is being processed, in the context of this UC.

2.5.2.3.1.5 Necessity

The UC Partner should carefully assess whether the purposes for which it intends to use the Platform can be met with only anonymous or aggregated data collected from signals and/or 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).

2.5.2.3.1.6 Compatibility

The presumption of compatibility mentioned above applies to this UC, for research and statistical purposes.

Data minimisation

Dataset 12 [18] is intended to be purged of identifiers concerning the citizens submitting signals before its processing via the Platform. However, this dataset currently includes information on the content provided by citizens within free text fields in each signal, which may reveal further personal data on the submitting citizens. To this regard, UC Partner has noted that no text analysis is intended to be performed on these fields, such that any such personal data will not be further processed in UC #3 [17].

2.6 Communication, Exploitation, Standardisation, Roadmapping & Business Development (WP7)

2.6.1 Data Marketplace ("DMP")

The DMP is designed as a smart, user-based asset repository. PolicyCLOUD users will be able to use this component to share ready-to-use solutions and/or tools (including, e.g., data sources, analytics tools, policy models) with other users. [29]

The main legal, regulatory, ethical, and societal requirements identified for the DMP component are as follows:

- The DMP should incorporate adequate technical and organizational security measures, developed as a result of a dedicated security risk assessment targeting potential threats.
- Should personal data be included in data sources made available through the DMP, the PolicyCLOUD users uploading such data sources to the DMP, the users accessing such data sources and/or the PolicyCLOUD manager(s) must identify appropriate legal bases under the GDPR (and implement all associated compliance steps) to ensure that this making available of personal data can be lawfully carried out, and that the relevant data sources can be lawfully leveraged by accessing users.
- All potential controllers involved in the uploading and accessing of data sources containing personal data via the DMP (i.e., uploading and accessing PolicyCLOUD users, the PolicyCLOUD manager(s)) should be clear and honest with the relevant data subjects about their identities, the methods they will use to process personal data, and the purposes for which they are to process those personal data.
- T&Cs for the use of the DMP should be defined, to properly regulate the service relationship established between the PolicyCLOUD manager(s) and PolicyCLOUD users. These T&Cs must be accepted for use of the DMP to be allowed.
- A due diligence exercise must be performed, and a compliance declaration obtained from PolicyCLOUD users uploading assets to the DMP, with regards to the compliance of those assets with applicable legal, regulatory, ethical, and societal requirements.

The status of implementation of data management requirements concerning the DMP, as of the date of this Deliverable, and the next steps towards implementation of any missing requirements, is reported in the Updated WP7 Legal/Ethical Checklist (see Annex 23 – WP7 Legal/Ethical Checklist).

3 Conclusion

In this Deliverable, a final update on the implementation of relevant legal, regulatory, ethical, and societal concerns requirements previously identified as applicable to the PolicyCLOUD platform was presented in Section 2. Furthermore, in Annexes Annex 1 – Final WP2 Legal/Ethical Checklist to Annex 23 – WP7 Legal/Ethical Checklist, new and updated Legal/Ethical Checklists for each PolicyCLOUD WP, containing a consolidated set of legal and ethical requirements and details on the status of their implementation, were provided.

Concerning legal and regulatory concerns and requirements, the focus was kept, for practicality and to assure that requirements could be defined in a PolicyCLOUD user-agnostic manner, on the EU legal and regulatory framework, as opposed to local laws applicable in specific Member States or third countries.

Therefore, this Deliverable provides a final set of requirements which are understood and accepted by the Consortium as vital to ensure the ethical and legal soundness of PolicyCLOUD, as well as a description of the specific measures taken to address each requirement.

References

[1] Community Research and Development Information Service (CORDIS). Policy Management through technologies across the complete data lifecycle on cloud environments, https://cordis.europa.eu/project/id/870675, retrieved 2022-10-27.

[2] PolicyCLOUD. Making data-driven policy management a reality across Europe, https://policycloud.eu/, retrieved 2022-10-27.

[3] PolicyCLOUD. *D3.3 - PolicyCLOUD's Societal and Ethical Requirements & Guidelines.* Bettiol Alberto and Taborda Barata Martim. 2020.

[4] PolicyCLOUD. *D3.6 - PolicyCLOUD's Societal and Ethical Requirements & Guidelines – M 22.* Bettiol Alberto and Taborda Barata Martim. 2021.

[5] European Union Agency for Cybersecurity. *EUCS – Cloud Services Scheme*, https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme, retrieved 2022-10-27.

[6] EGI Foundation. *EGI Pay4Use VO Service Level Agreement*, https://documents.egi.eu/public/RetrieveFile?docid=3667&filename=EGI%20VO%20P4U%20SLA_PolicyCLOUD_FINAL.pdf&version=1, retrieved 2022-10-27.

[7] EGI Foundation, *EGI Pay4Use VO Operational Level Agreement*, https://documents.egi.eu/public/RetrieveFile?docid=3667&filename=EGI%20VO%20P4U%20OLA%20RECAS_BARI_PolicyCLOUD_FINAL.pdf&version=1, retrieved 2022-10-27.

[8] EGI Foundation. *Data Processing Agreement*, https://documents.egi.eu/public/RetrieveFile?docid=3745&filename=Data_processing_agreement_EGI_Cloud_Compute.v3.pdf&version=1, retrieved 2022-10-27.

[9] European Data Protection Board, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf, retrieved 2022-10-27.

[10] EGI Foundation, *EGI Policies and Procedures,* https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers, retrieved 2022-10-27.

[11] EGI Foundation, *Policy on the Processing of Personal Data,* https://documents.egi.eu/public/RetrieveFile?docid=2732&filename=EGI-SPG-Privacy-V2.pdf&version=3, retrieved 2022-10-27.

[12] EGI Foundation, *Technical and organizational measures (TOM),* https://documents.egi.eu/public/RetrieveFile?docid=3737&filename=TOM-Overview_EGI_Foundation.pdf&version=1, retrieved 2022-10-27.

[13] PolicyCLOUD. *D3.7 – Cloud Infrastructure Incentives Management and Data Governance: Design and Open Specification 3*. Sanguino Maria Angeles. 2022.

[14] PolicyCLOUD. *D4.5 – Reusable Models & Analytical Tools: Design and Open Specification 3.* Biran Ofer. 2022.

[15] PolicyCLOUD. *D5.6 - Cross-sector Policy Lifecycle Management: Design & Open Specification 3.* Baroni Samuele. 2022.

[16] PolicyCLOUD. *D8.1 – POPD-Requirement No. 1.* Bettiol Alberto and Taborda Barata Martim. 2021.

[17] PolicyCLOUD. *D6.11 – Use Case Scenarios Definition & Design.* Sancho Javier. 2022.

[18] PolicyCLOUD. *D1.4 – Data Management Plan M24.* Munné Ricard. 2021.

[19] Global Terrorism Database. *Homepage*. https://www.start.umd.edu/gtd/, retrieved 2022-10-27.

[20] RAND. *RAND Database of Worldwide Terrorism Incidents*. https://www.rand.org/nsrd/projects/terrorism-incidents.html, retrieved 2022-10-27.

[21] Global Terrorism Database. *End User License Agreement with University of Maryland*. https://www.start.umd.edu/gtd/terms-of-use/, retrieved 2022-10-27.

[22] RAND. *Download the Database*. https://www.rand.org/nsrd/projects/terrorism-incidents/download.html, retrieved 2022-10-27.

[23] Twitter. *Twitter Terms of Service.* https://twitter.com/en/tos#intlTerms, retrieved 2022-10-27.

[24] Twitter. *Twitter Privacy Policy*, https://twitter.com/en/privacy, retrieved 2022-10-27.

[25] 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 2022-10-27.

[26] 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 2022-10-27.

[27] European Data Protection Supervisor, *A Preliminary Opinion on data protection and scientific research.* https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf, retrieved 2022-10-27.

[28] Call Sofia. *Information on the Processing of Personal Data.* https://call.sofia.bg/bg/Account/PrivacyPolicy, retrieved 2022-10-27.

[29] PolicyCLOUD. *D7.4 – Data Marketplace: Design and Open Specification.* Mavrogiorgou Argyro. 2021.

[30] statsmodels v0.13.2. *User Guide: Time Series analysis (tsa).* https://www.statsmodels.org/stable/tsa.html, retrieved 2022-10-27.

[31] Facebook. *Prophet*. https://facebook.github.io/prophet/, retrieved 2022-10-27.

[32] PeerJ. *Forecasting at scale*. J. Taylor Sean and Letham Benjamin. 2017. https://peerj.com/preprints/3190v2/, retrieved 2022-10-27.

[33] Proceedings of the 9th Python in Science Conference. *Statsmodels: Econometric and Statistical Modeling with Python*. Seabold Skipper and Perktold Josef. 2010. https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf, retrieved 2022-10-27.

[34] Proceedings of the International AAAI Conference on Web and Social Media. *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. Hutto C. and Gilbert Eric. 2014. https://ojs.aaai.org/index.php/ICWSM/article/view/14550, retrieved 2022-10-27.

[35] arXiv e-prints. *Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification*. Rietzler Alexander, Stabinger Sebastian, Opitz Paul and Engl Stefan. 2019. https://arxiv.org/abs/1908.11860, retrieved 2022-10-27.

[36] arXiv e-prints. *Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems*. Kiritchenko Svetlana and M. Mohammad Saif. 2018. https://arxiv.org/abs/1805.04508, retrieved 2022-10-27.

[37] arXiv e-prints. *Assessing Demographic Bias in Named Entity Recognition*. Mishra Shubhanshu, He Sijun and Belli Luca. 2020. https://arxiv.org/abs/2008.03415, retrieved 2022-10-27.

[38] *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.* Honnibal, M., & Montani, I. 2017.

[39] Linguistic Data Consortium. *OntoNotes Release 5.0*. Weischedel Ralph et al. 2013. https://catalog.ldc.upenn.edu/LDC2013T19, retrieved 2022-10-27.

[40] arXiv e-prints. *Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods.* Zhao Jieyu et al. 2018. https://arxiv.org/abs/1804.06876, retrieved 2022-10-27.

Annex 1 – Final WP2 Legal/Ethical Checklist

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
1	An information security management system, covering at least the operational units, locations and processes for providing the Platform, must be defined, implemented, maintained and continually improved.	This control was deemed relevant only for a potential commercialisation phase for the platform (i.e., not relevant to the Project itself).	N/A
2	A risk assessment about the accumulation of responsibilities or tasks on roles or individuals, regarding the provision of the Platform must be conducted. This risk assessment must cover at least the following areas, as far as these are applicable to the provision of the Platform and are in the area of responsibility of the Consortium: • Administration of rights profiles, approval and assignment of access and access authorisations. • Development, testing and release of changes. • Operation of the system components.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to the Project.	Not implemented. No special risk assessment was requested or carried out specifically for the Project, but - as documented in the OLA between EGI and the IaaS provider - the IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers).
3	The mitigating measures defined in the risk assessment must be implemented, privileging separation of duties, unless impossible for organisational or technical	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	Not implemented. No special risk assessment was conducted for PolicyCLOUD, and thus this control could not be implemented. Nevertheless, EGI has implemented several relevant technical/organisational

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	reasons, in which case the measures shall include the monitoring of activities to detect unauthorised or unintended changes as well as misuse and the subsequent appropriate actions.		security measures (https://documents.egi.eu/document/3737). While EGI has not entered into a data processing agreement with the IaaS provider for the Project, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control.
4	Information security requirements and considerations must be included in the management of PolicyCLOUD.	ICTLC	Implemented. Implemented through the management of this WP2 Legal/Ethical Checklist.
5	Risk assessments must be conducted on the entire perimeter of the Platform.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	Not implemented. No special risk assessment was conducted for PolicyCLOUD, and thus this control could not be implemented. Nevertheless, EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737). The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers). EGI further operates a risk management process as part of its information management system, which is certified against ISO 9001:2015

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			and ISO/IEC 20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/).
6	The results of risk assessments must be made available to relevant stakeholders.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	**Not implemented.** No special risk assessment was conducted for PolicyCLOUD, and thus this control could not be implemented. Nevertheless, EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737). The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers). EGI further operates a risk management process as part of its information management system, which is certified against ISO 9001:2015 and ISO/IEC 20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/).
7	Risk assessments must be revised at least after each major change that may affect the security of the Platform.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to the Project.	**Not implemented.** See Control 6.
8	Risks must be prioritized according to their criticality.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	**Not implemented.** See Control 6.
9	A plan must be implemented to treat risks according to their priority level by reducing or	EGI Input on security risk assessments which may have been conducted by the IaaS	**Not implemented.** See Control 6.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	avoiding them through security controls, by sharing them, or by retaining them, so that each risk level is reduced to a threshold which the risk owner(s) deems acceptable (Residual Risk).	provider regarding the infrastructure relevant to PolicyCLOUD.	
10	The risk treatment plan must be made available to relevant stakeholders.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	Not implemented. See Control 6.
11	If PolicyCLOUD users will share risks, the shared risks must be associated to Complementary Customer Controls (“**CCCs**”) and described in the user documentation.	ICTLC	Implemented. Interim Repository developed to mitigate risk-sharing with PolicyCLOUD users (once a dataset it is ingested, it becomes fully within the PolicyCLOUD security perimeter). Collection from relevant Partners of information on bias and trade-off management (for analytic functions) and bias, privacy and data protection and authorisation management (for datasets), to address the additional input parameters added for registration concerning built-in analytic functions and datasets relevant to approved UC scenarios. Insertion of disclaimer in PDT-PME User Handbook advising PolicyCLOUD users of the possibility of false positive and negative results and margin of error, and to use critical judgment when interpreting results obtained via PolicyCLOUD in the context of their policymaking activities. There is the possibility for

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			PolicyCLOUD to be integrated with external datasets and third-party infrastructure and networks after commercialisation, at which point shared risks may become more relevant.
12	The risk treatment plan must be revised every time risk assessments are revised.	EGI Input on security risk assessments which may have been conducted by the IaaS provider regarding the infrastructure relevant to PolicyCLOUD.	Not implemented. See Control 6.
13	Information security-sensitive positions relevant to PolicyCLOUD must be classified according to their level of risk, including positions related to IT administration and to the provisioning of the cloud service in the production environment, and all positions with access to data or system components.	EGI OKS Ensure there is an appropriate identification of personnel with administrative access to PolicyCLOUD platform backend components, and with access to data stored on the platform.	Partially implemented. EGI Using EGI Check-In, EGI can control who has access to PolicyCLOUD's Virtual Organisation, which defines who has access to OpenStack (and, therefore, who can create or destroy virtual machines for PolicyCLOUD). EGI Check-In also provides roles with administrative privileges, allowing for the approval or rejection of applications to join PolicyCLOUD's VO (e.g., VO manager). New roles can be set up to meet PolicyCLOUD's needs. Virtual machines are created by the different Partners, and creators are responsible for controlling access to their machines (and any services hosted on such machines). Access to the Platform's underlying cloud infrastructure is permitted only to identified and authorised personnel, including appropriately trained system administrators. OKS

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			OKS is responsible for the Implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. OKS also supports SOF in the implementation of their UC. No sensitive data have been available to OKS in connection with these responsibilities.
14	Each Partner must include in its employment contracts or on a dedicated code of conduct or ethics an overarching agreement from internal and external employees to act ethically in their professional duties.	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. **UBI / UPRC** This has been implemented at UBI and UPRC (at an organisation level). **EGI** EGI's Terms of Employment, as well as its Information Security and ICT policies, contain sections that cover required and forbidden behaviour when processing personal data, or when using EGI resources. **IBM** This has been implemented at IBM (IBM team is indeed bound to act ethically in professional duties). **MAG** Maggioli requires all its employees to act according to a code of ethics to which they accept to be bound at the signature of the contract. **OKS** Relevant clauses to this end exist in the contracts signed by OKS in relation to PolicyCLOUD.
15	Each Partner must ensure that all internal and external employees are required by their employment terms and conditions to comply with all applicable information	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. See Control 14.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	security policies and procedures.		
16	Each Partner must ensure that the employment terms for all internal and external employees include a non-disclosure provision, which shall cover any information that has been obtained or generated as part of PolicyCLOUD, even if anonymised and decontextualized.	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. See Control 14.
17	Each Partner must ensure that all employees have completed a relevant security awareness and training program defined for them.	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. **UBI / UPRC** This has been implemented at UBI and UPRC (at an organisation level). **EGI** All EGI employees are trained, at least, at an ISO 27001 Foundation level. Security awareness and training is managed as part of EGI's information management system process. **IBM** This has been implemented at IBM. **MAG** Maggioli Group routinely performs tests on the employees on this requirement. **OKS** OKS' IT department has provided instructions to the relevant staff in this respect.
18	Each Partner must apply a specific procedure to revoke the access rights and process appropriately the accounts and assets of internal and external employees if their	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. **UBI / UPRC** This has been implemented at UBI and UPRC (at an organization level). **EGI** EGI has a dedicated procedure to

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	employment is terminated or changed (e.g., where the employee no longer works with PolicyCLOUD).		manage leaving employees, which is documented in EGI's information management system process. **IBM** This has been implemented at IBM. **MAG** This has been implemented at Maggioli. **OKS** OKS' IT department is following specific procedures which meet the requirements of this control.
19	Each Partner must ensure that non-disclosure or confidentiality agreements are agreed with internal employees, external service providers and suppliers.	UBI EGI UPRC IBM LXS MAG OKS	Partially implemented. **UBI / UPRC** This has been implemented at UBI and UPRC (at an organisation level). **EGI** This has been implemented at EGI (at an organisation level), as documented in EGI's Terms of Employment ("Confidentiality" section) and EGI's Information Security and ICT policies. **IBM** This has been implemented at IBM (as a standard organizational measure). **MAG** Maggioli Group includes confidentiality / non-disclosure agreements in contracts signed both by employees and external collaborators. **OKS** Relevant clauses to this end exist in the contracts signed by OKS in relation to PolicyCLOUD.
20	An inventory of assets must be maintained. For each asset, the information needed to apply the PolicyCLOUD's	EGI Confirm whether IaaS provider maintains an asset inventory relevant to the infrastructure provided to PolicyCLOUD.	Partially implemented. **EGI** This has not been implemented (lack of information provided on this control).

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	risk management procedure must be recorded.	ICTLC	ICTLC The assets relevant to the Platform (infrastructure, microservices and components, analytic functions, datasets) are reported in Section 7 of D2.7.
21	Technical and organizational measures to ensure acceptable use and safe handling of assets must be implemented.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Implemented. EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered into a data processing agreement with the IaaS provider for the Project, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control.
22	Technical and organizational measures to ensure adequate commissioning and decommissioning of hardware that is used to provide the Platform in the production environment must be implemented. These must include (in the case of decommissioning) measures to ensure the complete and permanent deletion of the data or the proper destruction of the media.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
23	An asset classification schema that reflects for each asset the protection needs of the information it processes, stores, or transmits must be defined.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
24	When applicable, all assets must be labelled according to their classification in the asset classification scheme.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
25	Security perimeters must be defined in the buildings and premises related to the provision of the Platform.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
26	At least two security areas must be defined, with one covering all buildings and premises and one covering sensitive activities such as the buildings and premises hosting the information system to produce the Platform.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
27	A set of security requirements for each security area must be defined and communicated.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
28	Technical and organizational measures to ensure adequate physical access control to the security areas must be implemented, requiring at least one authentication factor for accessing any non-public area. Derogations in case of	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	emergency must be defined.		
29	A warning must be displayed at the entrance of all non-public perimeters concerning the limits and access conditions to these perimeters.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
30	Security perimeters must be protected with security measures to detect and prevent unauthorised access in a timely manner so that it does not compromise the information security of the Platform.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
31	Technical and organizational measures concerning the protection of equipment must be implemented, covering at least the following aspects: • Protecting power and communications cabling from interception, interference or damage. • Protecting equipment during maintenance operations. • Protecting equipment holding data during transport.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
32	Encryption must be used on the removable media and the backup media intended to move between security areas according to	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	the sensitivity of the data stored on the media.		
33	A set of security requirements related to external and environmental threats must be documented and communicated, addressing the following risks in accordance with the applicable legal and contractual requirements: • Faults in planning. • Unauthorised access. • Insufficient surveillance. • Insufficient air-conditioning. • Fire and smoke. • Water. • Power failure. • Air ventilation and filtration.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Not implemented Lack of information provided on this control.
34	Technical and organizational measures to plan for capacities and resources (personnel and IT resources) must be implemented, which shall include forecasting future capacity requirements to identify usage trends and manage system overload.	EGI LXS	Partially implemented. EGI EGI operates a capacity management process as part of its information management system, which is certified against ISO 9001:2015 and ISO/IEC 20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/).
35	Requirements must be included in contractual agreements with PolicyCLOUD users (or their organisations) regarding the provision of the Platform in case of capacity bottlenecks or personnel and IT resources outages.	ICTLC	Implemented. Addressed in the latest revisions to the PDT-PME and DMP T&Cs.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
36	Technical and organizational measures to ensure the monitoring of provisioning and de-provisioning of the services linked to the Platform, to ensure compliance with the contractual agreements with PolicyCLOUD users (or their organisations), must be implemented.	This control was deemed relevant only for a potential commercialisation phase for the Platform (i.e., not relevant to PolicyCLOUD itself).	N/A
37	PolicyCLOUD users must be able to control and monitor the allocation of the system resources assigned to them, if these corresponding cloud capabilities are exposed to those users.	UBI EGI UPRC LXS	Partially implemented. **UBI** This is not implemented at a Platform user-level, as it was deemed out of scope for PolicyCLOUD. **EGI** The OpenStack Horizon dashboard, when used to manage access to system resources, allows for easy control and monitoring of the allocation of resources assigned to PolicyCLOUD. **UPRC** Monitoring activities are conducted at a Project-level through the OpenStack dashboard provided by EGI.
38	Technical and organizational measures to protect the Platform systems and users from malware must be implemented, covering at least the following aspects: • Use of system-specific protection mechanisms. • Operating protection programs on system	UPRC OKS EGI LXS	Partially implemented. **UPRC** Appropriate anti-malware technical and organisational measures, meeting the mentioned requirements, have been applied to the DMP. **OKS** OKS is responsible for the implementation of PME, which is a component of the PDT and written with Angular. Furthermore, the Keycloak system secures PDT and PME

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	components under the responsibility of the Consortium that are used to provide the Platform in the production environment. - Operation of protection programs for employees' terminal equipment.		with user authentication and authorization and from CSRF attacks. **EGI** The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
39	Malware protection must be deployed, if technically feasible, on all systems that support the delivery of the Platform in the production environment.	EGI LXS	Partially implemented. **EGI** The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
40	Technical and organizational measures to ensure adequate data backup and recovery must be implemented.	EGI LXS	**EGI** As mentioned in the SLA (https://documents.egi.eu/document/3667), backups in EGI's federated cloud are not available

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			by default. The provision of backups should have been discussed and included in the SLA during its negotiation. Despite this, extra capacity (provided by a separate IaaS provider) has been provisioned at the request of the Consortium, but not centralised backup system is in place.
41	Technical and organizational measures to monitor the execution of data backups must be implemented.	EGI LXS	See Control 40.
42	Restore procedures must be assessed.	EGI LXS	See Control 40.
43	Backup data must be transferred to a remote location or transported on backup media to a remote location.	EGI LXS	See Control 40.
44	Where backup data is transmitted to a remote location via a network, the transmission of the data must take place in an encrypted form that corresponds to the state-of-the-art.	EGI LXS	See Control 40.
45	Technical and organizational measures to ensure the logging and monitoring of events on system components must be implemented.	IBM EGI	Implemented. **IBM** The logging and monitoring of events on system components is ensured through the implemented logging service. **EGI** The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
46	Technical and organizational measures to ensure the secure handling of derived data must be implemented.	IBM Cloud service derived data includes log data containing records of who used the service, at what times, which functions, types of data involved and so on. It can also include information about the numbers of authorized users and their identities. It can also include any configuration or customization data, where the cloud service has such configuration and customization capabilities.	Implemented. Measures to ensure the secure handling of derived data, inherent to the implemented logging service, are in place which meet the requirements of this control.
47	Log data must be monitored to identify events that might lead to security incidents.	IBM	Not implemented. This control is out-of-scope for PolicyCLOUD, and thus has not been implemented in the development environment used; however, in a production environment (commercialisation phase) the use of Grafana Cloud Logs will address this control.
48	Identified events must be reported for timely assessment and remediation.	IBM	Implemented. This control is addressed as an outcome of the ELK stack (GRAFANA KIBANA…).
49	All log data must be stored in an integrity-protected and aggregated form that allow its centralized evaluation.	IBM	Implemented. This is implemented through the storage of activation logs in the system's CouchDB.
50	Log data must be deleted when it is no longer required for the purpose	IBM	Not implemented. This control has not been implemented in the development environment used

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	for which they were collected.		for PolicyCLOUD, since retaining logs in this environment is useful to ease debugging (by using older logs for comparison purposes); however, in a production environment (commercialisation phase) it can be easily configured.
51	Communication between the assets to be logged and the logging servers must be authenticated and protected in integrity and confidentiality.	IBM	Not implemented. This control is not applicable to PolicyCLOUD.
52	Log data generated must allow an unambiguous identification of user accesses at the PolicyCLOUD user level to support analysis in the event of an incident.	IBM	Not implemented. This control has not been implemented in the development environment used for PolicyCLOUD, to facilitate its use by the Partners and bearing in mind that the ultimate production environment may be different from that used for development purposes; however, in a production environment (commercialisation phase) it can be set up.
53	Only authorized users may be allowed access to system components used for logging and monitoring under their responsibility.	IBM	Implemented. Access to system components used for logging and monitoring is restricted to a limited number of authorised users.
54	The system components for logging and monitoring must be monitored, and failures must be automatically detected and reported.	IBM	Implemented. This control is addressed as an outcome of the ELK stack (GRAFANA KIBANA…).
55	Technical and organizational measures to ensure the timely identification and addressing of vulnerabilities in the system components used	IBM EGI UPRC LXS	Partially implemented. IBM This control is out-of-scope for IBM's contribution to PolicyCLOUD, and thus has not been implemented in the development environment used;

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	to provide the Platform must be implemented.		however, in a production environment (commercialisation phase) the use of Grafana Cloud Logs will address this control. **EGI** The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents). **UPRC** Regarding the Cloud Gateways component, the "swagger-stats" monitoring tool and the Traefik load balancer have been implemented.
56	A scoring system for the assessment of vulnerabilities, which includes at least “critical” and “high” classes of vulnerabilities, must be implemented.	IBM EGI	Partially implemented. **IBM** This control is out-of-scope for PolicyCLOUD, and thus has not been implemented in the development environment used. **EGI** The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
57	A publicly and easily accessible online register of known vulnerabilities that affect the Platform and assets provided which PolicyCLOUD users must install or operate under their own responsibility must be published. This online register indicates at least the following information for every vulnerability: • A presentation of the vulnerability following an industry-accepted scoring system. • A description of the remediation options for that vulnerability. • Information on the availability of updates or patches for that vulnerability. • Information about the remediation or deployment of patches or updates by the Consortium or the PolicyCLOUD user, including detailed instructions for operations to be	EGI LSX	Partially implemented. **EGI** The EGI SVG (Security Vulnerability Group, see https://go.egi.eu/svg) maintains a repository of vulnerabilities that threaten the EGI infrastructure, for which an advisory has been produced: https://advisories.egi.eu/.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	performed by the user.		
58	Tests to detect known vulnerabilities on the system components used to provide the Platform must be conducted, ideally on a regular basis.	UBI UPRC EGI IBM LXS	Partially implemented. UBI All platform components under UBI responsibility are regularly assessed against known vulnerabilities. UPRC All components under UPRC's responsibility are regularly tested against known vulnerabilities. Stress-tests have also been applied. EGI The EGI SVG (Security Vulnerability Group, see https://go.egi.eu/svg) maintains a repository of vulnerabilities that threaten the EGI infrastructure, for which an advisory has been produced: https://advisories.egi.eu/. At the same time, this control has not been implemented on the part of the IaaS provider (lack of information provided on this control). IBM This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
59	All the system components that are used to provide the Platform must be hardened, according to accepted industry standards.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UPRC IBM LXS OKS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents). **UPRC** All components under UPRC responsibility (Data Cleaning, Cloud Gateways, Enhanced Interoperability and the DMP) have been hardened in accordance with these requirements. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase). **OKS** As mentioned in control 38, PME follows the industry standards and the Angular components have been implemented in PolicyCloud.
60	PolicyCLOUD user data must be segregated, stored and processed on shared virtual and physical resources to ensure the confidentiality and integrity of this data.	LXS UBI	Partially implemented. **UBI** UBI stores and processes PolicyCLOUD user data on shared virtual resources which are segregated.
61	Role and rights policies and procedures must be implemented for controlling access to information resources, in which at least the following aspects are covered: • Parameters to be considered for making access control decisions.	UBI EGI UPRC IBM LXS OKS	Partially implemented. **UBI** This is confirmed - access control mechanisms and user management tools meeting these requirements (e.g., Lightweight Directory Access Protocol/LDAP directories and additional mechanisms/tools implemented under the ISO/IEC 27001 standard) are in place at UBI (at an organisation level).

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	• Granting and modifying access rights based on the "least-privilege" principle and on the "need-to-know" principle. • Use of a role-based mechanism for the assignment of access rights. • Segregation of duties between managing, approving and assigning access rights. • Dedicated rules for users with privileged access. • Requirements for the approval and documentation of the management of access rights.		**EGI** EGI has implemented several relevant technical/organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. **UPRC** UPRC has implemented relevant technical/organisational security measures at the level of its organisation, which meet the requirements of this control. **IBM** Attribute-based access control (ABAC) has been implemented by IBM for the Cloud gateway. **OKS** OKS provides its source code to PolicyCLOUD, allowing it to be covered by the access control mechanisms applicable to the GitLab solution used.
62	Technical and organizational measures to ensure the adequate managing of accounts must be implemented, in which at least the following aspects are covered: • Assignment of unique usernames. • Definition of the distinct types of accounts supported, and assignment of access control parameters and roles to be	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** Access control mechanisms and user management tools meeting these requirements (e.g., Lightweight Directory Access Protocol/LDAP directories and additional mechanisms/tools implemented under the ISO/IEC 27001 standard) are in place at UBI (at an organisation level). **EGI** The EGI Check-In solution used in PolicyCLOUD complies with the REFEDS Research and Scholarship (https://refeds.org/category/research-and-scholarship) entity category and the Sirtfi67890960 framework

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	considered for each type. - Events leading to blocking and revoking accounts.		(https://refeds.org/wp-content/uploads/2016/01/Sirtfi-1.0.pdf). It is thus reasonable to maintain that this solution ensures sufficient attribute release, as well as operational security, incident response, and traceability for identity federations that support REFEDS Research and Scholarship and Sirtfi. OKS OKS provides its source code to PolicyCLOUD, allowing it to be covered by the account management mechanisms applicable to the GitLab solution used. UPRC Unique IDs have not been assigned to individual Data Marketplace users for increasing security and minimizing the risk of users' identification. Roles for the different types of supported accounts (i.e., user [verified via e-mail or unverified] and administrator) are defined. No events which may lead to the blocking or revocation of accounts have been identified since this has not been implemented in the context of the PolicyCloud Data Marketplace, as this has not been a technical requirement.
63	Technical and organizational measures must be implemented to ensure the adequate managing of personal user accounts and access rights to internal and external employees that comply with the role and rights	UBI EGI OKS UPRC LXS	Partially implemented. UBI Access control mechanisms and user management tools meeting these requirements (e.g., Lightweight Directory Access Protocol/LDAP directories and additional mechanisms/tools implemented under the ISO/IEC

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	concept and with the policies for managing accounts.		27001 standard) are in place at UBI (at an organisation level). **EGI** EGI service users are bound to EGI's policies and procedures (https://confluence.egi.eu/display/EGIPP/Information+for+End+Users), which bind the users to compliance with these requirements. They must accept and agree to abide by the e-Infrastructure Acceptable Use Policy (https://documents.egi.eu/document/3600). **OKS** OKS provides its source code to PolicyCLOUD, allowing it to be covered by the account management mechanisms applicable to the GitLab solution used. **UPRC** Technical and organisational measures to ensure adequate managing of personal user accounts and access rights, in line with the requirements of this control, have been implemented for the DMP.
64	Technical and organizational measures must be implemented to ensure the adequate managing of non-personal shared accounts and associated access rights that comply with the role and rights concept and with the policies for managing accounts.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** Access control mechanisms and user management tools meeting these requirements (e.g., Lightweight Directory Access Protocol/LDAP directories and additional mechanisms/tools implemented under the ISO/IEC 27001 standard) are in place at UBI (at an organisation level). **OKS** OKS is responsible for the Implementation of the PME, which is a component of the PDT,

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			which in turn is protected under the Keycloak system. **EGI** As EGI Check-In only allows the creation of personal accounts, this has not been implemented. **UPRC** As there are no non-personal shared accounts in the DMP, this has not been implemented.
65	Technical and organizational measures must be implemented to ensure the adequate managing of technical accounts and associated access rights to system components involved in the operation of the Platform that comply with the role and rights concept and with the policies for managing accounts.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** Access control mechanisms and user management tools meeting these requirements (e.g., Lightweight Directory Access Protocol/LDAP directories and additional mechanisms/tools implemented under the ISO/IEC 27001 standard) are in place at UBI (at an organisation level). **OKS** OKS is responsible for the Implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **EGI** As EGI Check-In only allows the creation of personal accounts, this has not been implemented. **UPRC** As there are no technical accounts in the DMP, this has not been implemented.
66	An automated mechanism to block user accounts after a certain period must be implemented.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** EGI Check-In will deactivate users accounts for which their membership to the corresponding Virtual

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			Organisation is not manually renewed. **OKS** OKS is responsible for the Implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **UPRC** The DMP and Cloud Gateways components have been integrated with the Keycloack system to ensure compliance with this control.
67	An automated mechanism to block user accounts after a certain number of failed authentication attempts must be implemented.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** EGI Check-In is a proxy service, connecting federated identity providers with EGI service providers - the responsibility for implementation of user account blocking mechanisms is upon such identity providers. Check-In has been integrated into PolicyCLOUD services (PDT, ongoing for the DMP). **OKS** OKS is responsible for the Implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **UPRC** The DMP and Cloud Gateways components have been integrated with the Keycloack system to ensure compliance with this control.
68	Technical and organizational measures must be implemented to	UBI EGI	Implemented. **UBI**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	ensure the adequate granting, updating, and revoking to a user account of access rights to resources of the information system of the Platform, and these procedures shall be compliant with the role and rights concept and with the policies for managing access rights.		This is implemented in PolicyCLOUD through the Keycloak system. **EGI** EGI service users are bound to EGI's policies and procedures (https://confluence.egi.eu/display/EGIPP/Information+for+End+Users), which bind the provider to compliance with these requirements. They must accept and agree to abide by the e-Infrastructure Acceptable Use Policy (https://documents.egi.eu/document/3600).
69	Shared accounts under the responsibility of the Consortium shall be assigned only to internal or external employees.	UBI EGI IBM UPRC LXS OKS	Partially implemented. **UBI** This is implemented in the Project through the Keycloak system. **EGI** As EGI Check-In only allows the creation of personal accounts, this has not been implemented. **IBM** This has not been implemented at IBM. **UPRC** As no shared accounts are used for any components under UPRC's responsibility, this has not been implemented by UPRC. **OKS** OKS is responsible for the Implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system.
70	Technical and organizational measures must be implemented to ensure the adequate provisioning of authentication	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **OKS**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	mechanisms, covering at least the following aspects: • The selection of mechanisms suitable for every type of account and each level of risk. • The protection of credentials used by the authentication mechanism. • The generation and distribution of credentials for new accounts. • Rules for the renewal of credentials, including periodic renewals, renewals in case of loss or compromise. • Rules on the required strength of credentials, together with mechanisms to communicate and enforce the rules.		OKS is responsible for the implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **EGI** The EGI Check-In solution used in the Project complies with the REFEDS Research and Scholarship (https://refeds.org/category/research-and-scholarship) entity category and the Sirtfi framework (https://refeds.org/wp-content/uploads/2016/01/Sirtfi-1.0.pdf). It is thus reasonable to maintain that this solution ensures sufficient attribute release, as well as operational security, incident response, and traceability for identity federations that support REFEDS Research and Scholarship and Sirtfi. **UPRC** Technical and organisational measures to ensure adequate provisioning of authentication mechanisms, in line with the requirements of this control, have been implemented for the DMP.
71	All authentication mechanisms must include a mechanism to block an account after a predefined number of unsuccessful attempts.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** EGI Check-In is a proxy service, connecting federated identity providers with EGI service providers - the responsibility for implementation of user account blocking mechanisms is upon

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			such identity providers. Check-In has been integrated into PolicyCLOUD services (PDT, ongoing for the DMP). **OKS** OKS is responsible for the implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **UPRC** The DMP and Cloud Gateways components have been integrated with the Keycloack system to ensure compliance with this control.
72	Technical and organizational measures must be implemented to ensure the adequate management of credentials, including at least: • Non-reuse of credentials. • Trade-offs between entropy and ability to memorize. • Recommendations for renewal of passwords. • Rules on storage of passwords.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** The EGI Check-In solution used in PolicyCLOUD complies with the REFEDS Research and Scholarship (https://refeds.org/category/research-and-scholarship) entity category and the Sirtfi framework (https://refeds.org/wp-content/uploads/2016/01/Sirtfi-1.0.pdf). Thus, this solution ensures sufficient attribute release, as well as operational security, incident response, and traceability for identity federations that support REFEDS Research and Scholarship and Sirtfi. **OKS** OKS is responsible for the implementation of the PME, which is a component of the PDT,

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			which in turn is protected under the Keycloak system. **UPRC** The DMP and Cloud Gateways components have been integrated with the Keycloack system to ensure compliance with this control.
73	Passwords must be only stored using cryptographically strong hash functions.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** The EGI Check-In solution used in PolicyCLOUD complies with the REFEDS Research and Scholarship (https://refeds.org/category/research-and-scholarship) entity category and the Sirtfi framework (https://refeds.org/wp-content/uploads/2016/01/Sirtfi-1.0.pdf). Thus, this solution ensures sufficient attribute release, as well as operational security, incident response, and traceability for identity federations that support REFEDS Research and Scholarship and Sirtfi. **OKS** OKS is responsible for the implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **UPRC** This is implemented for the DMP.
74	Sufficient partitioning measures between the information system providing the Platform and	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure	Partially implemented. **EGI** EGI provides access to cloud resources via Virtual

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	other information systems must be implemented.	provided to PolicyCLOUD (and specify measures implemented). UBI End-user level. LXS Data repository. UPRC OKS	Organisations (https://confluence.egi.eu/display/EGIG/Virtual+organisation), which restrict access to a specified selection of information systems to a selected group of users. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **UPRC** UPRC is responsible for the PolicyCloud Data Marketplace which is protected under the Keycloak system. **OKS** OKS is responsible for the implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system.
75	Suitable measures for partitioning between PolicyCLOUD users must be implemented.	UBI EGI OKS UPRC LXS	Partially implemented. **UBI** This is implemented in PolicyCLOUD through the Keycloak system. **EGI** When the cloud is accessed directly via the OpenStack Horizon Dashboard, each PolicyCLOUD user can see the others' resources. However, when the cloud is accessed via the PaaS Orchestrator, this control is fully implemented. To fully comply with this control, access should be only granted via the PaaS Orchestrator. Additionally, Virtual Organisations can contain groups with diverse levels of trust, and access to services can be filtered based on group

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			membership to meet the project demands. **OKS** OKS is responsible for the implementation of the PME, which is a component of the PDT, which in turn is protected under the Keycloak system. **UPRC** The DMP and Cloud Gateways components have been integrated with the Keycloack system to ensure compliance with this control.
76	Technical and organizational measures must be implemented to ensure adequate encryption and key management, in which at least the following aspects are covered: - Usage of strong encryption procedures and secure network protocols. - Requirements for the secure generation, storage, archiving, retrieval, distribution, withdrawal and deletion of the keys. - Consideration of relevant legal and regulatory obligations and requirements.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UBI (user data) LXS (datasets)	Partially implemented. **EGI** EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered into a data processing agreement with the IaaS provider for PolicyCLOUD, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control. **UBI** All platform components under UBI responsibility rely on hypertext transfer protocol secure ("HTTPS") and/or data

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			encryption measures meeting these requirements.
77	Strong encryption mechanisms for the transmission of data over public networks must be implemented.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UBI (user data) LXS (datasets)	Partially implemented. EGI EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered into a data processing agreement with the IaaS provider for the Project, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control. UBI All platform components under UBI responsibility rely on HTTPS and/or data encryption measures meeting these requirements.
78	Technical and organizational measures must be implemented to ensure the adequate encryption of data during storage.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UBI (user data) LXS (datasets)	Partially implemented. EGI This has not been implemented (lack of information provided on this control). UBI All platform components under UBI responsibility rely on HTTPS and/or data encryption measures meeting these requirements.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
79	Technical and organizational measures must be implemented to ensure adequate and secure key management, including at least the following aspects: • Generation of keys for different cryptographic systems and applications. • Issuing and obtaining public-key certificates. • Provisioning and activation of the keys. • Secure storage of keys including description of how authorised users get access. • Changing or updating cryptographic keys including policies defining under which conditions and in which manner the changes and/or updates are to be realised. • Handling of compromised keys. • Withdrawal and deletion of keys.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UBI (user data) LXS (datasets)	Partially implemented. EGI EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered a DPA with the IaaS provider for the Project, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control. UBI Key management is addressed by UBI internally, through mechanisms which address these requirements.
80	Technical and organizational measures must be implemented to ensure the adequate and prompt detection and response to network-	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented).	Partially implemented. EGI The National Research and Education Networks ("**NRENs**") and the members of the EGI federation are monitoring their

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	based attacks and to ensure the protection of information and information processing systems.	LXS UPRC	network. EGI's Computer Security Incident Response Team is coordinating the incident response. EGI federation members are bound by EGI's Security Incident Response Policy (https://confluence.egi.eu/display/EGIPP/Security+Incident+Response+Policy), as well as other EGI policies relevant to providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind them to compliance with these requirements.
81	Specific security requirements to connect within the PolicyCLOUD network must be implemented, defining at least: • When the security zones are to be separated and when the PolicyCLOUD users are to be logically or physically segregated. • What communication relationships and what network and application protocols are permitted in each case. • How the data traffic for administration and monitoring are segregated	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	from each other at the network level. • What internal, cross-location communication is permitted. • What cross-network communication is allowed.		
82	Trusted and untrusted networks must be distinguished, based on a risk assessment.	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
83	Trusted and untrusted networks must be separated into different security zones for internal and external network areas (and DMZ, if applicable).	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
84	Both physical and virtualized network environments must be designed and configured to restrict and monitor the connection to trusted or untrusted networks according to defined security requirements.	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
85	Each network perimeter must be controlled by security gateways.	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
86	Separate networks must be defined and implemented for the administrative	EGI LXS	EGI This has not been implemented (lack of information provided on this control).

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	management of the infrastructure and the operation of management consoles.		
87	Networks for administration must be logically/physically separated from PolicyCLOUD users' networks.	EGI LXS	EGI This has not been implemented (lack of information provided on this control).
88	Networks used to migrate or create virtual machines must be physically and logically segregated.	EGI LXS	EGI This has not been implemented (lack of information provided on this control).
89	Segregation mechanisms at network level between the data traffic of different PolicyCLOUD users must be implemented.	EGI LXS	EGI This has not been implemented (lack of information provided on this control).
90	Up to date all documentation of the logical structure of the network used to provision or operate the PolicyCLOUD must be maintained. This documentation must cover, at least, how the subnets are allocated, how the network is zoned and segmented, how it connects with third-party and public networks, and the geographical locations in which data are stored.	EGI LXS MAG	Partially implemented. EGI This has not been implemented (lack of information provided on this control). MAG Regarding the structure and architecture of the PDT and PME Layers, this has been done by Maggioli through the writing and submission of different Deliverables on the PolicyCloud Platform.
91	The confidentiality of data must be ensured by suitable procedures when offering functions for software-defined networking ("SDN").	EGI LXS	EGI This has not been implemented (lack of information provided on this control).

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
92	The functionality of SDN functions must be validated before providing new SDN features to PolicyCLOUD users or modifying existing SDN features.	EGI LXS	EGI This has not been implemented (lack of information provided on this control).
93	Technical and organizational measures must be implemented to protect the transmission of data against unauthorised interception, manipulation, copying, modification, redirection or destruction.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UBI (user data) LXS (datasets)	Partially implemented. EGI EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered a DPA with the IaaS provider for PolicyCLOUD, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control. UBI Data transmission security is addressed by UBI internally, through mechanisms which address these requirements.
94	The Platform must only be accessible by cloud services from other providers or PolicyCLOUD users' IT systems through documented inbound and outbound interfaces.	UBI	Implemented. All platform components under UBI responsibility have related self-documenting user interfaces ("UIs") and Swagger-documented APIs.
95	The interfaces shall be clearly documented for	UBI	Implemented. See Control 94.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	subject matter experts to understand how they can be used to retrieve the data.		
96	Communication on these interfaces shall use standardised communication protocols that ensure the confidentiality and integrity of the transmitted information according to its protection requirements.	UBI	Implemented. HTTPS and OpenAPI3 are the protocols used.
97	Communication over untrusted networks shall be encrypted.	EGI LXS	Partially implemented. EGI The IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers), which bind the provider to compliance with these requirements. In particular, the IaaS provider must abide by EGI's Service Operations Security Policy, Traceability and Logging Policy, and all other applicable EGI policies (https://confluence.egi.eu/display/EGIBG/SPG+Documents).
98	Contractual agreements applicable to PolicyCLOUD users (or their organisations) must include, at least, the following aspects concerning the termination of the contractual relationship: • Type, scope and format of the data the Platform	ICTLC	Implemented. This control is in place in the PDT/PME and DMP Terms and Conditions.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	provides to the user. • Delivery methods of the data to the user. • Definition of the timeframe, within which the Platform makes the data available to the user. • Definition of the point in time as of which the Platform makes the data inaccessible to the user and deletes the data. • Responsibilities and obligations to cooperate for the provision of the data.		
99	Technical and organizational measures must be implemented for deleting PolicyCLOUD users' data (including metadata and data stored in backups) upon termination of their contract in compliance with the relevant contractual agreements.	UBI LXS UPRC	Implemented. **UBI** A mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined). **LXS** Data manipulation abilities needed to allow for the effective exercise of GDPR Data Subject Rights (notably, the right to erasure) can be executed on the Platform's data repository (even if only manually, by system administrators). **UPRC** Technical and organisational measures to ensure adequate

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			data deletion mechanisms, in line with the requirements of this control, have been implemented for the DMP and for Cloud Gateways.
100	Technical and organizational measures must be implemented to ensure adequate change management of the IT systems supporting the PolicyCLOUD platform. Changes must be categorized and prioritized considering the potential security effects on the system components concerned and must be assessed and approved before deployment.	EGI Confirm whether IaaS provider has such measures in place regarding the infrastructure provided to PolicyCLOUD (and specify measures implemented). UPRC IBM LXS	Partially implemented. **EGI** EGI is operating a change management process as part of its information management system, which is certified against ISO 9001:2015 and ISO/IEC 20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/; https://confluence.egi.eu/display/EGIPP/EGI+Change+Management+CHM). **UPRC** Technical and organisational measures to ensure adequate change management, in line with the requirements of this control, have been implemented for Data Cleaning, Cloud Gateways, Enhanced Interoperability and the DMP. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
101	Roles and rights must be defined for the authorised personnel or system components who are allowed to update the Platform in the production environment.	UBI EGI UPRC LXS	Partially implemented. **UBI / UPRC** EGI has created specific user accounts with access rights to the production virtual machines (VMs), which have been assigned to UBI and UPRC. **EGI** EGI is operating a change management process as part of its information management system, which is certified against ISO 9001:2015 and ISO/IEC

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/; https://confluence.egi.eu/display/EGIPP/EGI+Change+Management+CHM).
102	All changes to the Platform in the production environment shall be logged and shall be traceable back to the individual or system component that initiated the change.	UBI EGI UPRC LXS	Partially implemented. **UBI / UPRC** EGI has provided system log generation and collection capabilities relevant to UBI components. These logs are stored by EGI. **EGI** EGI is operating a change management process as part of its information management system, which is certified against ISO 9001:2015 and ISO/IEC 20000-1:2018 (https://www.egi.eu/egi-foundation/certifications/; https://confluence.egi.eu/display/EGIPP/EGI+Change+Management+CHM).
103	Version control procedures must be implemented to track the dependencies of individual changes and to restore affected system components back to their previous state as a result of errors or identified vulnerabilities.	UBI UPRC LXS OKS	Partially implemented. **UBI / OKS** GitLab is used to ensure version control. **UPRC** Version control procedures, meeting the requirements of this control, are in place for all components under UPRC responsibility.
104	Technical and organizational measures must be implemented to ensure adequate and secure development of the Platform, which must consider information security from the earliest phases of design.	UBI UPRC LXS MAG OKS	Partially implemented. **UBI / OKS** Continuous integration pipelines are used in GitLab, which include security-related quality assurance aspects addressing these requirements. **UPRC** Technical and organisational measures to ensure adequate

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			and secure development, meeting the requirements of this control, are in place for all components under UPRC responsibility. **MAG** Maggioli offered its support in the implementation of these measures regarding the components of the platform under its responsibility.
105	A list of dependencies to hardware and software products used in the development of the Platform must be maintained.	UBI IBM UPRC LXS OKS	Partially implemented. **UBI** Docker-compose is used to list relevant dependencies regarding UBI components. **IBM** This has not been implemented at IBM. **UPRC** Lists of dependencies, meeting the requirements of this control, are maintained for all components under UPRC responsibility. **OKS** The source code and list(s) of dependencies for the PME are maintained in the Gitlab solution used.
106	The confidentiality and integrity of the source code must be protected at all stages of development.	UBI IBM UPRC LXS OKS	Partially implemented. **UBI / OKS** This is ensured by restricting GitLab access to authenticated users. **IBM** This has not been implemented at IBM. **UPRC** The source code for all components under UPRC responsibility is protected, as required by this control.
107	Version control must be used to keep a history of	UBI IBM	Partially implemented. **UBI / OKS**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	the changes in source code with an attribution of changes to individual developers.	UPRC LXS MAG OKS	This is ensured by restricting GitLab access to authenticated users. **IBM** This has been implemented at IBM as part of its standard development procedures. **UPRC** Version control mechanisms, meeting the requirements of this control, are in place for all components under UPRC responsibility. **MAG** Maggioli contributed by documenting all the details regarding the software components under its responsibility.
108	Production environments must be physically or logically separated from development, test or pre-production environments.	UBI UPRC IBM LXS OKS	Partially implemented. **UBI / UPRC / OKS** UBI, UPRC and OKS use their own development environment, which is physically and logically separated from the (EGI-provided) production environment. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
109	Data contained in the production environments must not be used in development, test or preproduction environments in order not to compromise their confidentiality.	UBI UPRC IBM LXS OKS	Partially implemented. **UBI / UPRC / OKS** UBI, UPRC and OKS only use test data in their development environment. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
110	Technical and organizational measures must be implemented to	UBI UPRC IBM	Partially implemented. **UBI / UPRC / OKS**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	ensure adequate checking of the Platform for vulnerabilities that may have been integrated into the Platform during the development process.	LXS OKS	Continuous integration pipelines are used in GitLab, which include security-related quality assurance aspects addressing these requirements. IBM This has not been implemented at IBM.
111	The procedures for identifying vulnerabilities must be integrated in the development process.	UBI UPRC IBM LXS OKS	Partially implemented. UBI / UPRC / OKS Continuous integration pipelines are used in GitLab, which uses Starboard to scan for and react to detected vulnerabilities. IBM This has not been implemented at IBM.
112	Technical and organizational measures must be implemented to ensure adequate controlling and monitoring of third parties whose products or services contribute to the provision of the Platform.	EGI	Implemented. EGI has implemented several relevant technical and organisational security measures (https://documents.egi.eu/document/3737), which meet the requirements of this control. While EGI has not entered a DPA with the IaaS provider for PolicyCLOUD, an OLA ("P4U OLA": https://documents.egi.eu/document/3667) has been executed between them, requiring the IaaS Provider to comply with the EGI Policy on the Processing of Personal Data (https://documents.egi.eu/document/2732), which includes requirements intended to cover the subject-matter of this control. As documented in the OLA between EGI and the IaaS provider, the IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			(https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers). EGI's services are monitored by its Service Monitoring solution, which produces availability and reliability reports. The security monitoring measures put in place by EGI's security coordination mechanisms ensure review of patch status and remediation of detected software vulnerabilities (https://www.egi.eu/service/security-coordination/).
113	A risk assessment of suppliers involved in the provision of the Platform must be performed and, where relevant, complementary requirements must be identified and implemented by those suppliers.	EGI	Not implemented. No special risk assessment was requested or conducted specifically for the Project, but - as documented in the OLA between EGI and the IaaS provider - the IaaS provider is bound by the EGI security policies and procedures applicable to EGI cloud providers (https://confluence.egi.eu/display/EGIPP/Information+for+Cloud+Providers).
114	A directory for controlling and monitoring the suppliers who contribute to the delivery of the Platform must be maintained.	EGI	Implemented. Under the OLA, the IaaS provider's service's availability and reliability is monitored via EGI's Service Monitoring solution, and regular reports are produced (which are shared with the Project, under the SLA).
115	Technical and organizational measures must be implemented to ensure adequate analysis, evaluation and treatment of identified violations and deviations by suppliers.	EGI	Implemented. See Control 114.
116	When a change in a third-party contributing to the	EGI	Implemented.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	delivery of the Platform affects its level of security, technical and organizational measures must be implemented to ensure that all PolicyCLOUD users can be informed of this without delay.		The SLA defines single points of contact, which are used to inform the Project of relevant changes whenever needed. Service downtime is documented in EGI's Configuration Database, and the Project (i.e., the Partners) can subscribe to them via EGI's Operations Portal.
117	Technical and organizational measures must be implemented to ensure a fast, effective and proper response to all known security incidents.	IBM (logging system) UBI EGI UPRC LXS OKS	Partially implemented. **IBM** This control is out of scope for IBM's contribution to PolicyCLOUD, and thus has not been implemented in the development environment used; however, in a production environment (commercialisation phase) it can be set up. **UBI / UPRC / OKS** Continuous integration pipelines are used in GitLab, which include security-related quality assurance aspects addressing these requirements. **EGI** This is implemented through EGI's Security Incident Response Policy (https://documents.egi.eu/document/2935) and CSIRT Security Incident Handling Procedure (https://confluence.egi.eu/display/EGIPP/SEC01+EGI+CSIRT+Security+Incident+Handling+Procedure).
118	Analyses of security incidents must be performed to identify recurrent or significant incidents and to identify the need for further protection.	UBI EGI IBM UPRC LXS OKS	Partially implemented. **UBI / OKS / UPRC** Continuous integration pipelines are used in GitLab, which include security-related quality assurance aspects addressing these requirements. **EGI**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
			This is implemented through EGI's Security Incident Response Policy (https://documents.egi.eu/document/2935) and CSIRT Security Incident Handling Procedure (https://confluence.egi.eu/display/EGIPP/SEC01+EGI+CSIRT+Security+Incident+Handling+Procedure). **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
119	All documents and evidence that provide details on security incidents must be archived.	UBI EGI UPRC IBM LXS OKS	Partially implemented. **UBI / OKS** All information relevant to identified security incidents and generated reports are stored on GitLab. **EGI** This is implemented through EGI's Security Incident Response Policy (https://documents.egi.eu/document/2935) and CSIRT Security Incident Handling Procedure (https://confluence.egi.eu/display/EGIPP/SEC01+EGI+CSIRT+Security+Incident+Handling+Procedure). **UPRC** Continuous integration pipelines are used in GitLab, which include security-related quality assurance aspects addressing these requirements. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
120	Security mechanisms and processes must be	UBI EGI	Partially implemented. **UBI**

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	implemented for protecting all the information related to security incidents in accordance with criticality levels and legal requirements in effect.	UPRC IBM LXS OKS	Relevant mechanisms and processes meeting these requirements have been implemented at UBI (at a company level), under the ISO/IEC 27001 standard. **EGI** This is implemented through EGI's Security Incident Response Policy (https://documents.egi.eu/document/2935) and CSIRT Security Incident Handling Procedure (https://confluence.egi.eu/display/EGIPP/SEC01+EGI+CSIRT+Security+Incident+Handling+Procedure). **UPRC / OKS** UPRC and OKS have implemented security mechanisms and processes at an organisational level which meet the requirements of this control. **IBM** This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
121	Technical and organizational measures must be implemented to ensure adequate business continuity and contingency management, which must include the need to perform a business impact analysis to determine the impact of any malfunction to the Platform or its infrastructure.	This control was deemed relevant only for a potential commercialisation phase for the platform (i.e., not relevant to PolicyCLOUD itself).	N/A
122	The CSP shall document the legal, regulatory, self-imposed and contractual requirements relevant to	ICTLC	Implemented. Mapped in the various Legal/Ethical Checklists, which are continuously updated and

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	the information security of the Platform.		monitored throughout the Project.
123	Guidelines and recommendations to assist PolicyCLOUD users with the secure configuration, installation, deployment, operation and maintenance of the Platform must be made publicly available.	MAG OKS UPRC UBI IBM LXS	Partially implemented. MAG Maggioli contributed to the writing of the User Manual, which informs users about the correct and safe use of the PolicyCLOUD platform. UBI / OKS Instructions are provided to PolicyCLOUD users in this respect in README files included in GitLab for the different components. UPRC This is implemented for the DMP: users can find this information in the DMP's User Manual, and in the "Frequently Asked Questions (FAQs)" section of the DMP website's "About" page. IBM This has not been implemented at IBM (as the PolicyCloud project is in the development phase).
124	Guidelines and recommendations applicable to the Platform in the version intended for productive use must be maintained.	MAG OKS UPRC UBI IBM LXS	Partially implemented. See Control 123.
125	Comprehensible and transparent information must be provided to PolicyCLOUD users on: • The relevant jurisdiction applicable to the provision of the Platform. • System component	ICTLC	Implemented. This control is in place in the PDT/PME and DMP Terms and Conditions and Privacy Policy.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	locations, including its subcontractors, where data is processed, stored and backed up.		
126	Sufficient information must be provided for PolicyCLOUD users' subject matter experts to determine to assess the suitability of the Platform's jurisdiction and locations from a legal and regulatory perspective.	ICTLC	Implemented. This control is in place in the PDT/PME and DMP Terms and Conditions and Privacy Policy.
127	PolicyCLOUD users must be offered error handling and logging mechanisms that allow them to obtain security-related information about the security status of the Platform as well as the data, services or functions it provides.	IBM (logging system) UBI UPRC LXS MAG	Partially implemented. UBI Error handling mechanisms have been implemented on UBI components, and a logging mechanism is currently available to platform administrators. IBM The logging service which has been implemented allows any platform component to create necessary logs. UPRC Error handling mechanisms have been implemented on UBI components, and a logging mechanism is currently available. MAG Maggioli offered organisational help in the implementation of different functions to guide users through the PolicyCloud platform or to allow the user to report errors.
128	A suitable session management system must be used that at least corresponds to the state-	UBI	Implemented. This control is implemented via Keycloak, which provides an adequate token-based session management system.

#	Description	Identified Owner(s) and Observations	Control Status + Owner Input
	of-the-art and is protected against known attacks.		
129	The following aspects must be ensured if PolicyCLOUD users operate virtual machines or containers with the Platform: • The user can restrict the selection of images of virtual machines or containers according to their specifications, so that they can only launch the images or containers released according to these restrictions. • In addition, these images provided by the Platform are hardened according to accepted industry standards.	EGI LXS	Partially implemented. EGI The list of relevant virtual machine images can be restricted by Virtual Organisation managers via AppDB (https://wiki.appdb.egi.eu/). See also EGI's Security Policy for the Endorsement and Operation of Virtual Machine Images (https://confluence.egi.eu/display/EGIPP/Security+Policy+for+the+Endorsement+and+Operation+of+Virtual+Machine+Images).

TABLE 1 – FINAL WP2 LEGAL/ETHICAL CHECKLIST

Annex 2 – Final WP3 Legal/Ethical Checklist

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
1	It is necessary to have a clear and specific framework in place with the IaaS provider, in which objectives, processes and results expected from the Platform are clearly set out, to specify the capabilities needed from the IaaS provider.	Implemented.	An SLA [6] and an OLA [7] with the IaaS service providers have been entered (EGI / ReCaS - Bari). No further action seems necessary here, given the contractual framework put in place between the Consortium, EGI and the IaaS provider.
2	The framework developed with the IaaS provider must define appropriate service levels 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.	Implemented.	See Control 1.
3	The environmental impact of the infrastructure necessary to support the functioning of the cloud-based system should be reduced to a minimum.	Implemented.	EGI has obtained confirmation from the IaaS provider that they closely monitor the power usage effectiveness of the data centre which supports the PolicyCLOUD infrastructure (to reduce inefficiencies which might cause unnecessary energy consumption, which in turn mitigates the data centre's environmental impact). They have further confirmed that energy consumption optimisation measures are in place, notably the ability to tune the cooling facilities used and adjust the amount of IT resources which are active at any given time. The implementation of a direct free-cooling solution, which could rely on natural fresh air of up to 23.º C in temperature

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			to regulate cooling (and which could provide greater power usage effectiveness than mechanical cooling), is planned.
4	Adequate technical and organisational security measures must be implemented in the components developed by WP3, developed because of a dedicated security risk assessment targeting potential threats generated from reliance on big data analysis. Specific measures aimed at ensuring data integrity (e.g., prevention of external attacks affecting the integrity of data used as input, the functioning of the analytics components, or the output generated by those components) should be considered in this assessment, as well as specific measures aimed at ensuring data confidentiality and availability (where input datasets may be stored on the Platform), and restrictions on reuse of personal data (e.g., hashing or encryption). The ability to detect and appropriately respond to security incidents, including personal data breaches, must also be considered. The technical and organisational security measures put in place by the IaaS provider in relation to the cloud infrastructure should sufficiently mitigate any relevant risks identified.	Partially implemented.	Matters related to security must be coordinated at the Project-level – European Union Agency for Cybersecurity (“**ENISA**”) (draft) “EUCS – Cloud Services Scheme” [5] has been identified as a relevant standard for PolicyCLOUD, and several controls have been extracted from this standard to serve as a security benchmark for the Project. Different Partners (including WP3 Partners) have been engaged to provide information on the status of implementation of these controls regarding the PolicyCLOUD components for which they are responsible. Where a given control has not been implemented and it is feasible to do so given the technical circumstances applicable to PolicyCLOUD, the Partner(s) responsible for such control has (have) been tasked with this implementation. The collection of information on the implementation of security controls and the monitoring of such implementation has primarily been conducted through the revision, consolidation, and implementation of the **WP2 Legal/Ethical Checklist**, together with all relevant Partners.
5	The Platform should be designed to facilitate the exercise of rights granted by the GDPR or other applicable data protection laws to the relevant data subjects (i.e., platform users). This may	Implemented.	UBI has confirmed that the data manipulation abilities described in "Data Subject Rights" (Table 3 of D3.6 [4]) can be executed on the WP3 Platform components which

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
	include measures such as setting up a dedicated e-mail inbox to receive requests for the exercise of these rights, as well as an internal workflow which allows for adequate management of these requests within the timeframe provided by the GDPR (as a rule, requests must be addressed within one month from receipt). Furthermore, the Platform should be designed to not create any relevant technical obstacles to the exercise of these rights.		store personal data, such as Platform user data (even if only manually, by system administrators).
6	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 Platform.	Implemented.	A Data Governance and Privacy Enforcement mechanism has been developed for the Platform. No further action seems necessary here, as this is addressed by the access control mechanisms implemented in connection with the activities conducted by WP3, and the terms and conditions developed for the PDT, DMP, etc. under WPs 5 and 7.
7	PolicyCLOUD shall keep personal data (of platform users) 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 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.	Implemented.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
8	A DPA shall be executed between PolicyCLOUD and the IaaS provider, including all the requirements defined by Art. 28 GDPR.	Implemented.	EGI has made template DPAs available, which may be used to complement the OLA and SLA with additional obligations upon EGI/the provider. The template DPA with EGI Foundation as a Processor relevant to Cloud Compute services [8] has been subject to a full assessment under Article 28 GDPR and relevant EDPB guidelines and found to be in overall alignment with Article 28 GDPR. Recommendations have been made by ICTLC to ensure its adequacy considering the personal data processing activities relevant to PolicyCLOUD and applicable best practices. EGI has provided a written confirmation of its role as processor (and, implicitly, of the IaaS provider as sub-processor) regarding the personal data used and processed by the Consortium on the resources provided by EGI to PolicyCLOUD (i.e., the provision of resources framed by the OLA [7] and the SLA [6]). They further confirmed that all such processing took place in accordance with the terms of EGI's relevant template DPA [8], EGI's relevant policies – notably their Policy on the Processing of Personal Data [10] – and a description of technical and

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			organisational security measures implemented by EGI to meet legal and contractual requirements when processing personal data [12]. As this confirms that the processing of personal data conducted by EGI and the IaaS provider within the context of PolicyCLOUD has abided by Art. 28 GDPR, it is reasonable to maintain that such a confirmation serves as an acceptable, albeit sub-optimal (from a compliance perspective) alternative to formally closing out a DPA for PolicyCLOUD.

TABLE 2 – FINAL WP3 LEGAL/ETHICAL CHECKLIST

Annex 3 – Final WP4 Legal/Ethical Checklist

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
1	The Platform's analytic components (including analytic-ingest components) must be covered by an appropriate initial and routine training and testing protocol or program. This protocol or program must aim to mitigate the risk that the components may skew knowledge obtained from data (output) in a biased or otherwise erroneous manner. It must also aim to ensure a reasonable degree of statistical accuracy for output generated (e.g., in the case of Enhanced Interoperability, for annotations and connections established).	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			platform) may be implemented, if deemed feasible and effective.
2	Adequate technical and organizational security measures must be implemented in the components developed by WP4, developed as a result of a dedicated security risk assessment targeting potential threats generated, in particular, from reliance on big data analysis. Specific measures aimed at ensuring data integrity (e.g., prevention of external attacks affecting the integrity of data used as input, the functioning of the analytics components, or the output generated by those components) should be considered in this assessment, as well as specific measures aimed at ensuring data confidentiality and availability (where input datasets may be stored on the Platform), and restrictions on reuse of personal data (e.g., hashing or encryption). The ability to detect and appropriately respond to security incidents, including personal data breaches, must also be considered.	Partially implemented.	Matters related to security must be coordinated at the Project-level - ENISA's (draft) "EUCS – Cloud Services Scheme" [5] has been identified as a relevant standard for PolicyCLOUD, and several controls have been extracted from this standard to serve as a security benchmark for the Project. Different Partners (including WP4 Partners) have been engaged to provide information on the status of implementation of these controls regarding the PolicyCLOUD components for which they are responsible. Where a given control has not been implemented and it is feasible to do so given the technical circumstances applicable to PolicyCLOUD, the Partner(s) responsible for such control has (have) been tasked with this implementation. The collection of information on the implementation of security controls and the monitoring of such implementation has primarily been conducted through the revision, consolidation, and implementation of the **WP2 Legal/Ethical Checklist**, together with all relevant Partners.
3	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) must be implemented. These logs should assist in ensuring traceability of automated decisions (i.e., explaining how an	Implemented.	A standard logging service has been implemented in the platform, as a centralized, PolicyCLOUD project-wide component (see Section 2.3.4). This logging mechanism has been further developed to ensure that it is also able to record events corresponding to the users'

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
	algorithm arrived at a given output from a given input), as well as the presentation of correlations established and their rationale to the end-user for confirmation. This information should allow end-users to validate correlations made to some extent.		invocation of Analytic Functions regarding specific Data Sources. The generated records and logs can be read in combination with the information provided on each invoked Analytic Function (as provided, e.g., in the function parameters, type, category, description and purpose parameters of the function's registration on the Platform) to allow for the tracking of specific actions and/or operations performed by an invoked Analytic Function on a given Data Source - in other words, to allow traceability of the process followed by an Analytic Function to convert the input (dataset) to the function's output.
4	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias and trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of Platform users. ICTLC has incorporated the final list of

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address has been set up on the PDT/PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during the Platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the Platform) may be implemented, if deemed feasible and effective.
5	Should the use of a selected data source be subject to restrictions (e.g., contractual terms, database copyright protection, database *sui generis* right protection, copyright protection), that data source should not be used without appropriate permissions from the relevant owners or rightsholders of that data source.	Implemented.	A requirement has been set for Data Owners seeking to register a Data Source to document measures taken to address applicable legal requirements, through adequate fields added to the registration APIs. Specific input parameters to be addressed by Data Owners require them to link information and documentation to confirm and/or demonstrate that the registration of the Data Source has been authorized by relevant rightsholders (or that such authorization is not required under the EU legal framework), to the Data Source upon registration. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			registration on the Platform (see Annex 26).
6	The Platform should be designed to facilitate the exercise of rights granted by the GDPR or other applicable data protection laws to the relevant data subjects (i.e., individuals whose personal data may be stored within cleaned and ingested datasets). This may include measures such as setting up a dedicated e-mail inbox to receive requests for the exercise of these rights, as well as an internal workflow which allows for adequate management of these requests within the timeframe provided by the GDPR (as a rule, requests must be addressed within one month from receipt). Furthermore, the Platform should be designed to not create any relevant technical obstacles to the exercise of these rights. Regarding personal data stored within cleaned or ingested datasets, it should be possible to manipulate those personal data as needed to respond to any of the requests described in the "Data Subject Rights" tab.	Implemented.	LXS has confirmed that the Platform's data repository allows for the execution of relevant data manipulation abilities (Table 3 of D3.6 [4]) needed to ensure the ability to address requests for the exercise of Data Subject Rights by individuals whose personal data may be included in a given Data Source (even if only manually, by system administrators).
7	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 Platform.	Implemented.	A Data Governance and Privacy Enforcement mechanism has been developed for the Platform. No further action seems necessary here, as this is addressed by the access control mechanisms implemented in connection with the activities conducted by WP3, and the terms and conditions developed for the PDT, DMP, etc., under WPs 5 and 7.
8	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	Implemented.	The PolicyCLOUD user is in control of the specific data points which will be cleaned or ingested and stored on the Platform – attribute

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
	processed. As such, for each purpose of processing connected to PolicyCLOUD, it shall identify the minimum amount of personal data needed to fulfil such purpose.		and data point filtering is part of the cleaning process. When registering a Data Source, the Data Owner can specify the attributes which are to be further processed as part of the Data source's parameters. This allows users to configure the Platform so that personal data is not unnecessarily collected - in other words, to either prevent or minimize the collection of personal data (e.g., by refraining from collecting relevant identifiers). A requirement has been set for Data Owners seeking to register a Data Source to document measures taken to address applicable legal requirements, through adequate fields added to the registration APIs. Specific input parameters to be addressed by registrants require them to link privacy and data protection management information and documentation to the data source upon registration. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26).
9	End-users and PolicyCLOUD must ensure that appropriate steps are taken to verify the accuracy of any personal data collected, and to maintain those personal data up to date over time.	Implemented.	The transformation and "pre-processing" of data, include activities such as: · Data validation, to ensure that the rest of the Platform components operates on clean, correct, and useful data. This also involves assessing the data against defined constraints (which

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			may be mandatory or optional), to provide further assurances as to the accuracy and consistency of data processed via the Platform. · Data cleaning, to correct or remove all data for which validation errors were raised during the validation process, including missing, irregular, unnecessary and inconsistent data. This also involves deleting or replacing data which is not aligned with the defined constraints. · Data verification, to check the data for accuracy and any remaining inconsistencies, after the validation and cleaning processes have been completed. Mandatory and optional data constraints can be defined to configure the parameters under which these data validation, cleaning and verification activities operate. This affords control to Data Owners and other PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This also allows the configuration of the Platform so that personal data is not unnecessarily collected or processed, thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform. Furthermore, various libraries are exploited in connection with the Platform's data validation, cleaning, and verification activities to provide greater assurances of data quality (including accuracy, completeness, and lack of errors). ICTLC has incorporated the final

#	Recommendation / Examples of activities concerned / further explanation	Implementation Status	Notes
			list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26).
10	End-users should also be advised of the possibility of false positive and negative correlations, so that they are incentivized to verify the validity of correlations made.	Implemented.	ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). Under WP5, disclaimers should be presented to PDT users to emphasise that output presented to them may not be 100% reflective of the underlying reality (due to the risk of false positive or negative correlations), thereby requiring users to exercise their own critical judgment in interpreting output and using it to inform policymaking. These are now inserted in the PDT-PME User Handbook. A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the Platform) may be implemented, if deemed feasible and effective.

TABLE 3 – FINAL WP4 LEGAL/ETHICAL CHECKLIST

Annex 4 – Updated WP5 Legal/Ethical Checklist

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Implemented.	For each of the UC scenarios involving the processing of personal data, a DPIA has been performed.
2	The Platform should incorporate adequate technical and organizational security measures, developed because of a dedicated security risk assessment targeting potential threats generated from reliance on big data analysis.	Partially implemented.	Matters related to security must be coordinated at the Project-level - ENISA's (draft) "EUCS – Cloud Services Scheme" [5] has been identified as a relevant standard for PolicyCLOUD, and several controls have been extracted from this standard to serve as a security benchmark for the Project. Different Partners (including WP5 Partners) have been engaged to provide information on the status of implementation of these controls regarding the PolicyCLOUD components for which they are responsible. Where a given control has not been implemented and it is feasible to do so given the technical circumstances applicable to PolicyCLOUD, the Partner(s) responsible for such control has (have) been tasked with this implementation. The collection of information on

			the implementation of security controls and the monitoring of such implementation has primarily been conducted through the revision, consolidation, and implementation of the **WP2 Legal/Ethical Checklist**, together with all relevant Partners.
3	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.	Implemented.	An appropriate legal basis for each purpose for which personal data on PolicyCLOUD users may be collected has been identified, under Art. 6 GDPR. All steps needed to properly implement the identified legal bases have been taken, depending on the legal bases selected, which in turn depends on the way user personal are processed via the PDT. The legal bases identified to this regard are: • Compliance with legal requirements, according with Art. 6, par. 1, let. c) GDPR. • The legitimate interest of both PolicyCLOUD and the organization of which the end user is part to the proper operation of the platform, according with Art. 6, par. 1, let. f) GDPR. • The legitimate interest of PolicyCLOUD and the organization using the Platform to protect the Platform and the information included in the same

			Platform, according with Art. 6, par. 1, let. f) GDPR. - The establishment, exercise, or defense of legal claims, according with Art. 6, par. 1, let. f) GDPR.
4	PolicyCLOUD should only manage personal data in ways that may be expected and not use such data in a way that may produce unjustified adverse effects on data subjects.	Implemented.	End users are aware of the ways in which their personal data may be processed, since a specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
5	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.	Implemented.	Mechanisms are in place to ensure that users can exercise their rights in relation to any personal data of theirs which may be collected during the use of the platform. More specifically, each data subject can exercise the following rights by sending a request in writing to PolicyCLOUD, to the extent permitted by applicable law: - To access. The data subject can obtain information relating to the processing of their personal data and a copy of such personal data. - To erase. The data subject can require the deletion of their personal data, to the

			extent permitted by law. • To object. The data subject can object to the processing of their personal data, on grounds relating to their situation. In cases of opposition to the processing of personal data, PolicyCLOUD reserves the right to assess the request, which will not be accepted if there are legitimate reasons to proceed with the processing that prevail over the freedoms, interests, and rights of the data subject. • To rectify. Where the data subject consider that their personal data is inaccurate or incomplete, they can require that such personal data be modified accordingly. • To restrict. The data subject can request the restriction of the processing of their personal data. Furthermore, should the data subject believe that the processing of their personal data is contrary to the legislation in force, they have the right to lodge a complaint to the competent data protection supervisory authority.
6	PolicyCLOUD and the end-users must be clear and honest with data	Implemented.	A specific privacy policy has been developed for the PDT,

	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.		to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
7	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 conducted in the context of PolicyCLOUD: a privacy policy.	Implemented.	See Control 6.
8	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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Implemented.	The data protection information notice is provided to the end users of the Platform using a two-layer approach: • The first layer is represented by a pop-up banner which appears to the end user when visiting the Platform. This pop-up banner includes a link to the second layer data protection information notice (i.e., the extended version of the data protection information notice). • The second layer is represented by the extended version of the data protection information notice, including all the elements required by

			Arts 13 and 14 GDPR. This extended version of the data protection information notice, the text of which is provided under D8.1. [16], is accessible to end users by clicking on a footer named "End Users Data Protection Information Notice" published on all the pages of the web environment on which the Platform operates.
9	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.	Implemented.	T&Cs for the use of the PDT and PME have been defined, to properly regulate the service relationship established between the PolicyCLOUD manager(s) and tool users (i.e., individual users, or organizations to which the individual users belong). These T&Cs will need to be accepted for the use of the PDT and PME to be allowed.
10	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.	Implemented.	See Control 9.
11	PolicyCLOUD should implement technical measures, including hashing and cryptography, to limit the possibility of repurposing personal data.	Implemented.	The PDT implements a HTTPS layer.
12	PolicyCLOUD shall only collect personal data, which is adequate, relevant, and limited to what is necessary in relation to the	Implemented.	Only the strict minimum amount of user personal data needed for the above listed purposes are collected.

	purposes for which they are processed. As such, for each purpose of processing connected to PolicyCLOUD, it shall identify the minimum amount of personal data needed to fulfil such purpose.		More specifically, the personal data of the end users of the platform processed by PolicyCLOUD are: 1. Name. 2. E-mail address. 3. Organization and role. 4. Username and password. 5. IP address. 6. Event logs related to the use of the Platform.
13	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.	Implemented.	Users are given the possibility to rectify any personal data they submit, or which is collected on them, in connection with use of the PDT, by sending a request in writing to PolicyCLOUD.
14	PolicyCLOUD shall implement technical and organizational 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.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection or processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further

			storage and processing via the Platform.
15	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 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.	Implemented.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
16	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.	Partially implemented.	See Control 2.
17	There should be clearly defined rules and specific channels on the	Partially implemented.	See Control 2.

	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.		
18	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 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 conducted, 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 organizational measures to reduce or eliminate the likelihood of recurrence.	Partially implemented.	See Control 2.
19	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	Implemented.	For each of the UC scenarios involving the processing of personal data, a DPIA has been performed.

	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.		
20	A security risk assessment, as part of an overall DPIA, should be conducted, to identify 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.	Implemented.	See Control 19.
21	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 organization to which the end-user belongs. These terms and conditions would need to be accepted for the use of the PDT to be allowed.	Implemented.	T&Cs for the use of the PDT and PME have been defined, to properly regulate the service relationship established between the PolicyCLOUD manager(s) and tool users (i.e., individual users, or organizations to which the individual users belong). These T&Cs will need to be accepted for the use of the PDT and PME to be allowed.

TABLE 4 – UPDATED WP5 LEGAL/ETHICAL CHECKLIST

Annex 5 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario A (Radicalization Incidents)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	In the context of UC #1, scenario A (Radicalization incidents) does not raise specific concerns with regards to personal data compliance. Indeed, for this scenario the relevant data sources identified in D1.4 [18] are dataset 7 (GTD) [19] and dataset 8 (RDWTI) [20], which include non-personal data and/or personal data of public domain (e.g., information regarding well known terrorists, victims of terrorist attacks, and/or any other person involved in terroristic events).
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely evaluate the analytics components of the platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and/or processing of personal data, such as by refraining from requiring the

			collection/processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented (to be balanced with EU Directive 2016/680).	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results

			generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 5.
6	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 authorization from the rights holder.	Implemented.	See Control 5.
7	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	Closed / Not Applicable.	In the context of UC #1, scenario A (Radicalization incidents) does not raise specific concerns with regards to personal data compliance. Indeed, for this scenario the

	data subjects to correct, complete or update their own personal data when needed (to be balanced with EU Directive 2016/680).		relevant data sources identified in D1.4 [18] are dataset 7 (GTD) [19] and dataset 8 (RDWTI) [20], which include non-personal data and/or personal data of public domain (e.g., information regarding well known terrorists, victims of terrorist attacks, and/or any other person involved in terroristic events).
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 7.
9	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 (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 7.
10	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 conducted in the context of the Project: a privacy policy (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 7.

11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Closed / Not Applicable.	See Control 7.
12	Effective anonymization or aggregation methods shall be implemented.	Closed / Not Applicable.	See Control 7.
13	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 7.

TABLE 5 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 1: SCENARIO A (RADICALIZATION INCIDENTS)

Annex 6 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario B (Radicalized Groups and Individuals)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	In the context of UC #1, scenario B (Radicalized Groups and Individuals) does not raise specific concerns with regards to personal data compliance. Indeed, for this scenario the relevant data sources identified in D1.4 [18] are dataset 7 (GTD) [19] and dataset 8 (RDWTI) [20], which include non-personal data and/or personal data of public domain (e.g., information regarding well known terrorists, victims of terrorist attacks, and/or any other person involved in terroristic events).
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the

			collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented (to be balanced with EU Directive 2016/680).	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on

			the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	End-users and PolicyCLOUD must ensure that appropriate steps are taken to verify the accuracy of any	Closed / Not Applicable.	See Control 1.

	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 (to be balanced with EU Directive 2016/680).		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 1.
9	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 (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 1.
10	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 conducted in the context of the Project: a privacy policy (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 1.
11	PolicyCLOUD shall make available the privacy policy on its cloud-based platform, with appropriate steps taken to make it available to the data subjects whose personal	Closed / Not Applicable.	See Control 1.

	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.		
12	Effective anonymization or aggregation methods shall be implemented.	Closed / Not Applicable.	See Control 1.
13	All relevant GDPR and EU Directive 2016/680 principles (e.g., lawfulness, legal basis, purpose limitation, etc.) shall be covered.	Closed / Not Applicable.	See Control 1.
14	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted (to be balanced with EU Directive 2016/680).	Closed / Not Applicable.	See Control 1.

TABLE 6 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 1: SCENARIO B (RADICALIZED GROUPS AND INDIVIDUALS)

Annex 7 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario C (Trend Analysis)

Control No.	Recommendation / Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 (to be balanced with EU Directive 2016/680).	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant

			identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented (to be balanced with EU Directive 2016/680).	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or

			biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information/documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	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	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities

	subjects to correct, complete or update their own personal data when needed (to be balanced with EU Directive 2016/680).		operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection or processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC (to be balanced with EU Directive 2016/680).	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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 (to be balanced with EU Directive 2016/680).	Implemented.	See Control 8.
10	PolicyCLOUD and the end-users, in relation to the processing activities which they may respectively	Implemented.	See Control 8.

	perform, as controllers, shall lay down a specific and easily accessible document which duly informs data subjects of the processing activities conducted in the context of PolicyCLOUD: a privacy policy (to be balanced with EU Directive 2016/680).		
11	PolicyCLOUD shall make 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.	Implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been

			implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR and EU Directive 2016/680 principles (e.g., lawfulness, legal basis, purpose limitation, etc.) shall be covered.	Implemented.	For this UC scenario, a DPIA has been performed.
14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Implemented.	See Control 13.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted (to be balanced with EU Directive 2016/680).	Implemented.	See Control 13.

TABLE 7 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 1: SCENARIO C (TREND ANALYSIS)

Annex 8 – Updated WP6 Legal/Ethical Checklist – UC # 1: Scenario D ([Near] Real-time Assessment of Online Propaganda)

Control No.	Recommendation / Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 (to be balanced with EU Directive 2016/680).	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data,

			such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented (to be balanced with EU Directive 2016/680).	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will

			be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	End-users and PolicyCLOUD must ensure that appropriate steps are	Implemented.	See Control 2.

	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 (to be balanced with EU Directive 2016/680).		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC (to be balanced with EU Directive 2016/680).	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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 (to be balanced with EU Directive 2016/680).	Implemented.	See Control 8.
10	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 conducted in the context of PolicyCLOUD: a privacy policy (to be balanced with EU Directive 2016/680).	Implemented.	See Control 8.
11	PolicyCLOUD shall make 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	Implemented.	See Control 8.

	PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.		
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR and EU Directive 2016/680 principles (e.g., lawfulness, legal basis, purpose limitation, etc.) shall be covered.	Implemented.	See Control 1.

14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Implemented.	See Control 1.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted (to be balanced with EU Directive 2016/680).	Implemented.	See Control 1.

TABLE 8 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 1: SCENARIO D ([NEAR] REAL-TIME ASSESSMENT OF ONLINE PROPAGANDA)

Annex 9 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario A.1 (Politika Price Point)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data

			to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic

			functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	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	Implemented.	See Control 2.

	correct, complete or update their own personal data when needed.		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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.	Implemented.	See Control 8.
10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope

			for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Closed/Not applicable.	No targeting of social media users is performed under this UC scenario.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Closed/Not applicable.	No minor data is processed under this UC scenario.

TABLE 9 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 2: SCENARIO A.1 (POLITIKA PRICE POINT)

Annex 10 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario A.2 (Price Evolution)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data

			to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic

			functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	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	Implemented.	See Control 2.

	correct, complete or update their own personal data when needed.		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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.	Implemented.	See Control 8.
10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope

			for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Closed/Not applicable.	No targeting of social media users is performed under this UC scenario.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Closed/Not applicable.	No minor data is processed under this UC scenario.

TABLE 10 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 2: SCENARIO A.2 (PRICE EVOLUTION)

Annex 11 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario B (Opinion on Social Networks)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data

			to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic

			functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	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	Implemented.	See Control 2.

	correct, complete or update their own personal data when needed.		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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.	Implemented.	See Control 8.
10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope

			for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Closed/Not applicable.	No targeting of social media users is performed under this UC scenario.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Closed/Not applicable.	No minor data is processed under this UC scenario.

TABLE 11 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 2: SCENARIO B (OPINION ON SOCIAL NETWORKS)

Annex 12 – Updated WP6 Legal/Ethical Checklist – UC # 2: Scenario C (Trend Analysis)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data

			to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic

			functions used on the Platform, as well as other errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Implemented.	Input parameters (*biasDoc*, *GDPRDoc* and *authDoc*) related to the data source registration APIs require registrants to link information and/or documentation to confirm and/or demonstrate that the registration of a data source has been authorised by relevant rightsholders.
5	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 authorization from the rights holder.	Implemented.	See Control 4.
6	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 authorization from the rights holder.	Implemented.	See Control 4.
7	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	Implemented.	See Control 2.

	correct, complete or update their own personal data when needed.		
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Implemented.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
9	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.	Implemented.	See Control 8.
10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Implemented.	Data retention period definition has been determined as out of scope

			for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The requirements of the EDPB Guidelines 08/2020 on the targeting of social media users (considering the databases to be used) shall be respected.	Closed/Not applicable.	No targeting of social media users is performed under this UC scenario.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Closed/Not applicable.	No minor data is processed under this UC scenario.

TABLE 12 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 2: SCENARIO C (TREND ANALYSIS)

Annex 13 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario A (Visualization)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to

			their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a

			specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Closed / Not applicable.	Since the end-user (i.e., Sofia Municipality) is also responsible for the management of the contact center, there are no contractual restrictions towards leveraging information obtained via the contact center for the purposes of the end-user. For the same reason, there are no intellectual property law limitations related to the use of the data source by the end user.
5	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
6	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
7	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.	Partially implemented.	Dataset 12 [18] is intended to be purged of identifiers concerning the citizens submitting signals before its processing via PolicyCLOUD. However, this dataset currently includes information on the content provided by citizens within free text fields

			in each signal, which may reveal further personal data on the submitting citizens. To this regard, the corresponding UC partner has noted that no text analysis is intended to be performed on these fields, such that any such personal data will not be further processed in the context of UC #3 [16].
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Partially implemented.	With regards to the data protection information notice published on the Sofia Municipality website [28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR.*"
9	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.	Partially implemented.	See Control 8.

10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Partially implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Partially implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Partially implemented.	See Control 7.
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The end-user should identify steps to assure itself of the reliability of the 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	Implemented.	See Control 3.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Partially implemented.	See Control 7.

TABLE 13 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 3: SCENARIO A (VISUALIZATION)

Annex 14 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario B (Predictive Analysis)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to

			their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other errors which may occur during platform use – in the future, a

			specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Closed / Not applicable.	Since the end-user (i.e., Sofia Municipality) is also responsible for the management of the contact center, there are no contractual restrictions towards leveraging information obtained via the contact center for the purposes of the end-user. For the same reason, there are no intellectual property law limitations related to the use of the data source by the end user.
5	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
6	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
7	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.	Partially implemented.	Dataset 12 [18] is intended to be purged of identifiers concerning the citizens submitting signals before its processing via PolicyCLOUD. However, this dataset currently includes information on the content provided by citizens within free text fields

			in each signal, which may reveal further personal data on the submitting citizens. To this regard, the corresponding UC partner has noted that no text analysis is intended to be performed on these fields, such that any such personal data will not be further processed in the context of UC #3 [16].
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Partially implemented.	With regards to the data protection information notice published on the Sofia Municipality website [28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR."*
9	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.	Partially implemented.	See Control 8.

10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Partially implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Partially implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Partially implemented.	See Control 7.
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The end-user should identify steps to assure itself of the reliability of the 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	Implemented.	See Control 3.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Partially implemented.	See Control 7.

TABLE 14 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 3: SCENARIO B (PREVENTIVE ANALYSIS)

Annex 15 – Updated WP6 Legal/Ethical Checklist – UC # 3: Scenario C (Environment and Air Quality Cross Analysis)

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	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 compliance with the GDPR and with other applicable data protection laws.	Implemented.	For this UC scenario, a DPIA has been performed.
2	Data source quality must be controlled, by ensuring that only reliable sources are used, and to routinely assess the analytics components of the Platform to ensure that they do not skew knowledge obtained from data in a biased manner.	Implemented.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary

			personal data to be removed from data sources prior to their further storage and processing via the Platform.
3	Any trade-off between requirements, principles or individual rights considered in AI system development should be properly documented.	Implemented.	The specific Consortium members responsible for each relevant tool have been engaged to provide information about the bias/trade-off management measures applicable (or to be applied, where relevant) to the Analytic Functions for which they are responsible. This has been achieved through two lists of questions regarding which they have been prompted to provide input. ICTLC has revised the questionnaire answers regarding all relevant Analytic Functions. Input parameters have been defined in the Analytic Function registration process for Analytic Owners to provide information about the bias and trade-off management measures applicable to the Analytic Function they are seeking to register, for the benefit of the Platform users. ICTLC has incorporated the final list of questions in a broader User Guide, developed to further clarify the information which should be provided by hypothetical users during analytic tool and dataset registration on the Platform (see Annex 26). A specific e-mail address will be set up on the PDT-PME for PolicyCLOUD users to report any potentially skewed or biased results generated by analytic functions used on the Platform, as well as other

			errors which may occur during platform use – in the future, a specific interface for error and bias reporting (e.g., a support form or reporting function set up on the platform) may be implemented, if deemed feasible and effective.
4	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 authorization from the data source owner.	Closed / Not applicable.	Since the end-user (i.e., Sofia Municipality) is also responsible for the management of the contact center, there are no contractual restrictions towards leveraging information obtained via the contact center for the purposes of the end-user. For the same reason, there are no intellectual property law limitations related to the use of the data source by the end user.
5	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
6	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 authorization from the rights holder.	Closed / Not applicable.	See Control 4.
7	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	Partially implemented.	Dataset 12 [18] is intended to be purged of identifiers concerning the citizens submitting signals before its processing via PolicyCLOUD. However, this dataset currently includes information

	correct, complete or update their own personal data when needed.		on the content provided by citizens within free text fields in each signal, which may reveal further personal data on the submitting citizens. To this regard, the corresponding UC partner has noted that no text analysis is intended to be performed on these fields, such that any such personal data will not be further processed in the context of UC #3 [16].
8	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, also considering the need to preserve the quality of information in cases where providing this information may have an impact on the effectiveness of the UC.	Partially implemented.	With regards to the data protection information notice published on the Sofia Municipality website [28][24], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR.*"
9	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,	Partially implemented.	See Control 8.

	the methods used to process personal data, and the purposes of processing.		
10	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 conducted in the context of PolicyCLOUD: a privacy policy.	Partially implemented.	See Control 8.
11	PolicyCLOUD shall make 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 PolicyCLOUD. End-users should likewise ensure that the above information is available to data subjects on public websites under their control.	Partially implemented.	See Control 8.
12	Effective anonymization or aggregation methods shall be implemented.	Partially implemented.	See Control 7.
13	All relevant GDPR principles (e.g., lawfulness, legal basis, purpose limitation, etc.) are covered.	Implemented.	See Control 1.
14	The end-user should identify steps to assure itself of the reliability of the 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	Implemented.	See Control 3.
15	The processing of minor data should be subjected to effective anonymization methods whenever feasible, even where individuals are targeted.	Partially implemented.	See Control 7.

TABLE 15 – UPDATED WP6 LEGAL/ETHICAL CHECKLIST – UC # 3: SCENARIO C (ENVIRONMENT AND AIR QUALITY CROSS ANALYSIS)

Annex 16 – DPIA – UC # 1: Scenario C (Trend Analysis)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#1 – Scenario C (Trend Analysis). Thanks to the output of Scenario C, policymakers will have an innovative tool to monitor social media for radicalization. Our stakeholders are concerned about the spreading of extreme political and religious ideas through media on which they have little to no control on. Thanks to this scenario, they will be able to identify clouds of keywords and monitor the evolution of the topic on social media.	===
3	DPO, if involved	Not applicable	===
4	What type of personal data are processed?	Dataset 4: relevant Twitter posts published by users, captured, and processed for subsequent analysis in UC #1.	===
5	Who are the recipients of personal data?	PolicyCLOUD's end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for

			revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Yes.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	Art. 6, par. 1, let. e) GDPR.
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and

			leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing	Yes.	See Control 12.

	and right to restriction of processing?		
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as medium, and thus acceptable.			

TABLE 16 – DPIA – UC # 1: SCENARIO C (TREND ANALYSIS)

Annex 17 – DPIA – UC # 1: Scenario D ([Near]real-time Assessment of Online Propaganda)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#1 – Scenario D ([Near]real-time Assessment f Online Propaganda). The policy maker can select the keywords of his/her interest and consult the different information linked to them on social media and online. This will allow for a near real-time assessment of the online instances of propaganda in various groups and social media that will enable the policy maker to have a much higher understanding of the situation.	===
3	DPO, if involved	Not applicable	===
4	What type of personal data are processed?	Dataset 4: relevant Twitter posts published by users, captured, and processed for subsequent analysis in UC #1.	===
5	Who are the recipients of personal data?	PolicyCLOUD’s end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the

			GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Yes.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	Art. 6, par. 1, let. e) GDPR.
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the

			specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of	Yes.	See Control 12.

	rectification, right of erasure, right to object to processing and right to restriction of processing?		
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as medium, and thus acceptable.			

TABLE 17 – DPIA – UC # 1: SCENARIO D ([NEAR]REAL-TIME ASSESSMENT OF ONLINE PROPAGANDA)

Annex 18 – DPIA – UC # 2: Scenario B (Opinion on Social Networks)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#2 – Scenario B (Opinion on social networks). Visualize the negative and positive opinions on social networks of the various products analyzed allowing an automatic and immediate response to the end user.	===
3	DPO, if involved	Not applicable	===
4	What type of personal data are processed?	Dataset 9 (Wine varieties and brands information from Twitter)	===
5	Who are the recipients of personal data?	PolicyCLOUD's end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is

			therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Yes.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	Art. 6, par. 1, let. e) GDPR.
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of

			relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing and right to restriction of processing?	Yes.	See Control 12.
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as medium, and thus acceptable.			

TABLE 18 – DPIA – UC # 2: SCENARIO B (OPINION ON SOCIAL NETWORKS)

Annex 19 – DPIA – UC # 2: Scenario C (Trend Analysis)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#2 – Scenario C (Trend Analysis). Analysis of the trends in the wine sector through the specialized websites of the sector.	===
3	DPO, if involved	Not applicable	===
4	What type of personal data are processed?	Dataset 9 (Wine varieties and brands information from Twitter)	===
5	Who are the recipients of personal data?	PolicyCLOUD's end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).

7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD’s Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Yes.	A specific privacy policy has been developed for the PDT, to provide written information to users as to how their personal data may be managed when using the PDT in a concise, transparent, intelligible, and easily accessible form, using clear and plain language, meeting all the requirements of Arts. 13 and 14 GDPR.
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	Art. 6, par. 1, let. e) GDPR.
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected (‘data minimization’)?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform’s data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data

			sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing and right to restriction of processing?	Yes.	See Control 12.
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as medium, and thus acceptable.			

TABLE 19 – DPIA – UC # 2: SCENARIO C (TREND ANALYSIS)

Annex 20 – DPIA – UC # 3: Scenario A (Visualization)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#3 – Scenario A (Visualization). Visualize the signals received via Sofia's Call Centre CallSofia related to 1) road infrastructure; 2) environment and air quality; 3) waste collection and waste disposal; 4) transport, and parking; 5) cleanliness of public spaces; 6) violation of public order. Provide a detailed analysis of their frequency over time and territorial distribution by categories / types, areas, districts, etc., to support and facilitate data-based municipal decision-making.	===
3	DPO, if involved	Not applicable.	===
4	What type of personal data are processed?	Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality [18].	===
5	Who are the recipients of personal data?	PolicyCLOUD's end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for

			such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Partially.	With regards to the data protection information notice published on the Sofia Municipality website [24][28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling,*

			monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR."
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	The processing of special categories of personal data is conducted for statistical purposes and according with Art. 25m of the Bulgarian Personal Data Protection Act, which states that "*Personal data originally collected for a different purpose may be processed for the purposes of the National Archive Funds, for scientific, for historical research or for statistical purposes. In such cases, the controller shall apply appropriate technical and organizational measures to safeguard the rights and freedoms of the data subject in accordance with Article 89 (1) of Regulation (EU) 2016/679."*".
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal

			data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing	Yes.	See Control 12.

	and right to restriction of processing?		
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as low, and thus acceptable.			

TABLE 20 – DPIA – UC # 3: SCENARIO A (VISUALIZATION)

Annex 21 – DPIA – UC # 3: Scenario 3 (Predictive Analysis)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#3 – Scenario B (Predictive Analysis). Using predictive analysis to predict a future outcome.	===
3	DPO, if involved	Not applicable.	===
4	What type of personal data are processed?	Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality [18].	===
5	Who are the recipients of personal data?	PolicyCLOUD's end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been

			implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Partially.	With regards to the data protection information notice published on the Sofia Municipality website [28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR.*"
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	The processing of special categories of personal data is conducted for statistical

			purposes and according with Art. 25m of the Bulgarian Personal Data Protection Act, which states that "*Personal data originally collected for a different purpose may be processed for the purposes of the National Archive Funds, for scientific, for historical research or for statistical purposes. In such cases, the controller shall apply appropriate technical and organizational measures to safeguard the rights and freedoms of the data subject in accordance with Article 89 (1) of Regulation (EU) 2016/679.*".
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.

12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing and right to restriction of processing?	Yes.	See Control 12.
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			
Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as low, and thus acceptable.			

TABLE 21 – DPIA – UC # 3: SCENARIO B (PREDICTIVE ANALYSIS)

Annex 22 – DPIA – UC # 3: Scenario C (Environment and Air Quality Cross-Analysis)

#	Question	Answer	Notes
1	Date	28/10/2022	===
2	Title and description of the application/project	UC#3 – Scenario C (Environment and Air Quality Cross-Analysis). Provide a cross-analysis of the datasets on environment and air quality received via Sofia’s call center Call Sofia and the additional data provided by Sofia’s air quality measurement stations to support and facilitate data-based municipal decisions in one of the key areas of urban development.	===
3	DPO, if involved	Not applicable.	===
4	What type of personal data are processed?	Dataset 12 (Sofia Municipality Signals), i.e., signals from citizens, coming through the Call Sofia contact center of the municipality [18].	===
5	Who are the recipients of personal data?	PolicyCLOUD’s end users.	===
6	Do you have a specified data retention period?	No.	Data retention period definition has been determined as out of scope for PolicyCLOUD, since it has been agreed that it is not practically feasible to define a retention period for personal data on Platform users shorter than the duration of PolicyCLOUD. After PolicyCLOUD is completed, this responsibility will be transferred to the organisation(s) responsible for the management of the Platform, which may be qualified as controller(s) for such personal data under the GDPR – these organisations

			will then be responsible for revising this retention period, to ensure compliance with the GDPR's principle of storage limitation. UBI has confirmed that a mechanism for deletion of Platform user personal data has been implemented via Keycloak. It is therefore possible for an administrator to easily delete personal data of Platform users, which can be used to enforce retention periods (once defined).
7	Indicate the assets through which the application/project processes personal data	PolicyCLOUD's Platform.	===
8	Does the application/project process data under rules approved in a code of conduct, under Article 40 GDPR?	No.	===
9	Are the purposes of processing described specified (clear and unambiguous), explicit (communicated in a clear and understandable manner to data subjects), legitimate (not forbidden by law) and coherent (accurately reflecting the real purposes for which data are processed)?	Partially.	With regards to the data protection information notice published on the Sofia Municipality website [28], it would be appropriate to integrate it with the script suggested in D8.1 [16]: "*The personal data provided to Sofia Municipality may be, after anonymization and/or aggregation, used for research, analytical, statistical, and policymaking purposes. More specifically, the data may be used for the development of public policies, through the entire lifecycle of policy management (therefore including policy modelling, monitoring, enforcing, simulation, analysis, and compliance). The legal basis of the processing is the*

			performance of a task conducted in the public interest or in the exercise of official authority, according with Art. 6(1)(e) GDPR."
10	Have suitable legal bases for each of the processing purposes been identified?	Yes.	The processing of special categories of personal data is conducted for statistical purposes and according with Art. 25m of the Bulgarian Personal Data Protection Act, which states that "*Personal data originally collected for a different purpose may be processed for the purposes of the National Archive Funds, for scientific, for historical research or for statistical purposes. In such cases, the controller shall apply appropriate technical and organizational measures to safeguard the rights and freedoms of the data subject in accordance with Article 89 (1) of Regulation (EU) 2016/679.*".
11	Is data processing adequate, relevant, and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?	Yes.	Mandatory and optional data constraints can be defined to configure the parameters under which the Platform's data validation, cleaning and verification activities operate. This affords control to PolicyCLOUD users over the specific data points of a data source to be registered and leveraged via the Platform. This, in turn, will allow users to configure the Platform so that personal data is not unnecessarily collected or processed (i.e., allowing users to prevent or minimise the collection and processing of personal data, such as by refraining from requiring the

			collection and processing of relevant identifiers), thereby allowing unnecessary personal data to be removed from data sources prior to their further storage and processing via the Platform.
12	Are data subjects able to exercise their right of access and right to portability?	Yes.	It has been confirmed that the abilities described in Table 3 of D3.6 [4] – i.e., technical abilities which the PolicyCLOUD platform should allow (either under individuals' autonomous control, or under the control of platform system administrators) to ensure that Data Subject Rights can be appropriately exercised – can be covered via the platform, at least through manual intervention by system administrators. It thus appears, as of the date of this Deliverable, that the cloud-based infrastructure on which the PolicyCLOUD platform is hosted does not present any relevant technical obstacles to the implementation of these abilities and, consequently, to the exercise of data subject rights.
13	Are data subjects able to exercise their right of rectification, right of erasure, right to object to processing and right to restriction of processing?	Yes.	See Control 12.
14	Is there a procedure in place to inform recipients of any requests for rectification, erasure, objection, or restriction of the processing of personal data?	Yes.	See Control 12.
Risk Analysis			

Based on the above analysis, the risk level for the rights and freedoms of data subjects can be qualified as low, and thus acceptable.

TABLE 22 – DPIA – UC # 3: SCENARIO C (ENVIROMENT AND AIR QUALITY CROSS-ANALYSIS)

Annex 23 – WP7 Legal/Ethical Checklist

Control No.	Recommendation /Examples of activities concerned / Further explanation	Implementation status	Notes
1	The DMP should incorporate adequate technical and organizational security measures, developed as a result of a dedicated security risk assessment targeting potential threats.	Partially implemented.	Matters related to security must be coordinated at the Project-level - ENISA's (draft) "EUCS – Cloud Services Scheme" [5] has been identified as a relevant standard for PolicyCLOUD, and several controls have been extracted from this standard to serve as a security benchmark for the Project. Different Partners (including WP7 Partners) have been engaged to provide information on the status of implementation of these controls regarding the PolicyCLOUD components for which they are responsible. Where a given control has not been implemented and it is feasible to do so given the technical circumstances applicable to PolicyCLOUD, the Partner(s) responsible for such control has (have) been tasked with this implementation. The collection of information on the implementation of security controls and the monitoring of such implementation has primarily been conducted through the revision, consolidation, and implementation of the **WP2 Legal/Ethical Checklist**, together with all relevant Partners.
2	Should non aggregated and/or anonymized personal data made available through the DMP, an appropriate legal basis to do so shall be identified, according with the GDPR.	Implemented.	Data processing is based on the legitimate interest.
3	It should be defined T&Cs for the use of the DMP, to properly regulate the service relationship established between PolicyCLOUD and the	Implemented.	Addressed in the latest revisions to the DMP T&Cs.

	end user or the organization to which the end-user belongs. These T&Cs would need to be accepted for the participation to the DMP to be allowed.		
4	A due diligence shall be performed, and a compliance declaration shall be obtained with regards to the ethical, legal, regulatory, and societal compliance of the datasets eventually uploaded to the DMP.	Implemented.	Addressed in the latest revisions to the DMP T&Cs.

TABLE 23 – UPDATED WP7 LEGAL/ETHICAL CHECKLIST

Annex 24 – User Guide for functions and data source registration

The User Guide for functions and data source registration reads as follows (as of the date of this Deliverable):

Dear User,

When registering functions or data sources on the PolicyCLOUD platform, you will be asked to provide information on several input parameters. The purpose of this Guide is to provide you with detailed guidance on the information that you should provide to address each input parameter.

(i) Registration of functions on the PolicyCLOUD platform

There are two types of functions that can be registered on the PolicyCLOUD platform:

i. **Analytics Ingest / Transformation Functions** – This category includes functions that are used to apply initial analytics and/or transformation on the data fusion path of data sources, after which they are stored in PolicyCLOUD backend.

ii. **Analytics Functions** – This category includes functions that are activated to perform an analysis on a specified data source (which has already been ingested / transformed) to provide analytics insights that can be further used in support of policy decisions.

When registering a function on the PolicyCLOUD platform, you will be asked to address the following input parameters:

Input Parameter	What you need to provide
a. name	Please indicate the name of the function that you aim to register on the PolicyCLOUD platform.
b. kind	Please specify the function implementation type (e.g., java8, image based, etc.).
c. filename	Please specify the filename of the JAR file containing the function code.
d. main	Please specify the function's main class (which is the parameter with which the platform user will specify the main function to be initially invoked).
e. image	Please note that this input parameter is **optional** and is only relevant if the "kind" specified above (let. b) is "image based." Should you wish to complete this input parameter, please specify the image name residing in the container registry service.
f. function parameters	Please specify the function parameters (Json).
g. type	Please specify whether you aim to register: 1. an Analytics Ingest / Transformation Function, **or** 2. an Analytics Function.

h. category	Please specify the category of the function that you aim to register (e.g., Data-Cleaning, Data-Interoperability, Situational-Knowledge, Opinion-Mining, Social-Dynamic).
i. description	Please provide a description of the function that you aim to register and its intended usage.
j. Expected input data schema/format	Please specify the data schema / format that you expect will be inputted in the function you aim to register.
k. purpose	Please specify the purpose for which your function can be lawfully used (this will determine the users who will be granted access to your function on the platform).
l. biasDoc	Please note that, to meet this input parameter, you should provide information on the specific measures that you have implemented to address and document the risk of biases inherent to the function that you aim to register. Topics that you will need to cover include: • whether you have assessed the possible limitations of your function in achieving fair/unbiased results prior to its registration on the PolicyCLOUD Platform. • whether any potential biases that may be generated/reinforced using your function have been identified. • what is the likelihood that your function may produce biased/unfair results and who may be impacted the most by such biased results. • the measures implemented to mitigate the risk of generating (unfair) biased results. • whether your function is based on an AI model.
m. tradeoffsDoc	Please note that, to meet this input parameter, you will be asked to provide information on any trade-offs that you may have encountered when developing the function that you aim to register and on the decisions that you have taken in relation to any such trade-offs. Topics that you will need to cover include: • whether you have identified any competing interests and values relevant to the functioning of your function and if you have assessed whether there are any relevant trade-offs between those interests and values. • the methodology that you used to identify and quantify such trade-offs and the impact that such trade-offs may have on data subjects. • whether technical approaches that may be implemented to minimize trade-offs have been considered. • what considerations have been considered in striking a balance between the competing interests and values at stake.

TABLE 24 - INPUT PARAMETERS (FUNCTIONS)

(ii) Registration of Data Sources on the PolicyCLOUD platform

There are three types of Data Sources that can be registered on the PolicyCLOUD platform:

i. **Streaming** – This category includes data which is continuously streamed from one or more external sources (e.g., twitter) into the PolicyCLOUD data store.
ii. **Ingest-now** – This category includes external (or local) data which should be ingested into the PolicyCLOUD data store.
iii. **Existing** – This category includes data that is already existing in the PolicyCLOUD backend, without having been registered and then ingested through the PolicyCLOUD Data Acquisition and Analytics API. Registering this category of Data Source may result in transformations of the data and the modification of the Data Source which is stored in the PolicyCLOUD backend.

When registering a Data Source on the PolicyCLOUD platform, you will be asked to provide information on the following input parameters:

Input Parameter	What you need to provide
a. Name	Please indicate the name of the Data Source that you aim to register on the PolicyCLOUD platform.
b. Type	Please specify the type of Data Source that you aim to register, between: 1. Streaming. 2. Ingest-now. 3. Existing.
c. Description	Please provide a description of the Data Source that you aim to register. Be sure to clearly describe the origin of data collected, the population of subjects to which the data refers, and any other information needed for a user to understand whether the Data Source is representative of the population they may wish to target in the context of a given policymaking activity.
d. Source specification	Please provide information on source specification (e.g., streaming details, source data location, source URL, access method, credentials etc.).
e. Schema / metadata	Please describe the data schema of the Data Source that you aim to register on the PolicyCLOUD platform.
f. Permissions	Please specify by whom the Data Source can be used (whether the Data Source is public or whether it can be accessed by a specified user list) and for what purpose. Please also note that Analytics Functions which are to be applied to a Data Source should have been registered for purposes which are identical or compatible with the purposes identified for that Data Source.
g. Analytics Ingest Function + parameters if applicable	Please specify the parameters applicable to the Analytics Ingest Function (if any).
h. biasDoc	Please provide bias management documentation relevant to the Data Source in question. This refers to any assessments you have conducted concerning the risk that the Data Source may carry some form of inherent bias (e.g., in the collection or arrangement of data). This should include information on the bias detection methods applied to the Data Source, the specific biases identified, and the measures taken to address any such biases.

i. GDPRDoc	Please provide the privacy and data protection management documentation relevant to the Data Source in question. This documentation should indicate whether any personal data is included within the Data Source and, if so, the measures that you have taken to ensure that the Data Source can be ingested onto the PolicyCLOUD platform in compliance with the applicable data protection laws (namely, the GDPR and any specific local privacy/data protection laws).
j. authDoc	Please provide the relevant documentation to demonstrate that the registration of the Data Source has been authorized by relevant rightsholders or, alternatively, please explain why no such authorization is required under applicable laws (e.g., the Data Source is a fully public and open-source dataset).

TABLE 25 - INPUT PARAMETERS (DATA SOURCES)

***

We hope we have clarified any doubts which may exist around the registration of Analytic Function and Data Sources on the PolicyCLOUD platform. If you have any questions, please reach to us at: ________.