



Policy Cloud

Cloud for Data-Driven Policy Management

CLOUD FOR DATA-DRIVEN POLICY MANAGEMENT

Project Number: 870675

Start Date of Project: 01/01/2020

Duration: 36 months

D2.5 STATE OF THE ART & REQUIREMENTS ANALYSIS M22

Dissemination Level	PU
Due Date of Deliverable	31/10/2020
Actual Submission Date	08/11/2021
Work Package	WP2, Requirements, Architecture & Innovation
Task	T2.1 Requirements Elicitation & State of the Art Analysis
Type	Report
Approval Status	
Version	V1.0
Number of Pages	p.1 – p.190

Abstract: This document contains the analysis of the state of the art of the baseline technologies that will be used in the scope of the project. Moreover, the elicitation of the requirements coming both from the use cases and the technological partners is included.

The information in this document reflects only the author's views and the European Community 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 his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



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

Versioning and Contribution History

Version	Date	Reason	Author
0.0	01/09/2021	Initial ToC	LXS
0.1	01/09/2021	Initial input on Executive summary, introduction, methodology, summary and conclusions	LXS
0.2	01/10/2021	Update of requirements	ICCS/EGI
0.3	08/10/2021	Update of Legal and Ethical Framework	ICTLC
0.4	11/10/2021	Update of requirements	ATOS/IBM/LXS/UPRC
0.5	13/10/2021	Update of use case of SOFIA	OKYS/SOF
0.6	14/10/2021	Update of use case of LONDON	LON
0.7	20/10/2021	Update of use case of MAGIOLI	MAG
0.8	21/10/2021	Update of use case of ITA/SARGA	ITA/SARGA
0.9	28/10/2021	All input gathered. Finalizing the document	LXS
0.10	02/11/2021	Internal review	UPRC
0.11	02/11/2021	Internal review	UBI
1.0	08/11/2021	All comments addressed. Submitted version	LXS, all

Author List

Organisation	Name
LXS	Boyan Kolev, Jose María Zaragoza, Patricio Martinez, Luis Miguel Garcia
IBM	Ofer Biran
ICCS	Kostas Moutselos, Marios Koniaris, Christos Pavlatos, Aggelos Kolaitis
ATOS	María Ángeles Sanguino, Jorge Montero, Ana Luiza Pontual, Miquel Milà, Ricard Munné
EGI	Giuseppe La Rocca
UPRC	Dimosthenis Kyriazis, Athanasios Kiourtis, George Manias, Argyro Mavrogiorgou
ITA	Rafael del Hoyo
UBI	Giannis Ledakis, Konstantinos Theodosiou, Konstantinos Oikonomou

OKS	Petya Nikolova, Christodoulos Maragkos, Stelios Betas, Konstantinos Nasias
MAG	Armend Duzha, Nikos Achilleopoulos
SOF	Iskra Yovkova, Ana Georgieva, Metodiya Tarlyovska
ITA	Vega Rodríguez
SARGA	Javier Sancho

Abbreviations and Acronyms

Abbreviation/Acronym	Definition
ABAC	Attribute-based access control
ALFA	Abbreviated Language For Authorization
API	Application Programming Interface
DBMS	Database Management Systems
BDAs	Big Data Analytics
EBPM	Evidence Based Policy Making
EC	European Commission
EIF	European Interoperability Framework
EOSC	European Open Science Cloud
HDFS	Hadoop Distributed File System
HTML	Hypertext Markup Language
IaaS	Infrastructure-as-a-Service
ICT	Information and Communication Technologies
IMI	Information Modelling and Interoperability
JSON	JavaScript Object Notation
KPI	Key Performance Indicators
MIDAS	Meaningful Integration of Data, Analytics and Services
NIF	National Improvement Framework
MLP	Neuro-Linguistic Programming
PaaS	Platform-as-a-Service
PHPM	Public Health Policy Model
PHPDM	PHP decision making
RACI	Responsibility assignment matrix
RDF	Resource Description Framework
SaaS	Service-as-a-Service
SQL	Structured Query Language
TCP	Transmission Control Protocol
VML	Vector Markup Language
XAMCL	extensible Access Control Markup Language
XML	extensible Markup Language
WP	Work Package

Contents

Versioning and Contribution History.....	3
Author List.....	3
Abbreviations and Acronyms	4
Executive Summary	15
1 Introduction.....	16
1.1 Method.....	17
1.2 Summary of Changes.....	20
2 Use case Requirements and Scenarios	21
2.1 UC#1: Participatory policies against radicalization.....	21
2.1.1 Goals and Objectives	21
2.1.2 Description of Scenarios	22
2.1.3 Stakeholder Requirements	27
2.2 UC#2: Intelligent policies for the development of agrifood industry.....	34
2.2.1 Goals and Objectives	34
2.2.2 Description of Scenarios	34
2.2.3 Stakeholder Requirements	45
2.3 UC#3 - Facilitating urban policy making and monitoring through crowdsourcing data analysis .	
.....	53
2.3.1 Goals and Objectives	53
2.3.2 Description of Scenarios	53
2.3.3 Stakeholder Requirements	60
2.4 UC#4 - Predictive analysis towards unemployment risks identification and policy making.....	61
2.4.1 Goals and Objectives	61
2.4.2 Description of Scenarios	61
2.4.3 Stakeholder Requirements	64
3 Use case datasets and data regulatory constraints.....	67
3.1 Dataset Specifications.....	67
3.1.1 UC#1 - Participatory policies against radicalization	67
3.1.2 UC#2 - Intelligent policies for the development of agrifood industry	71

3.1.3	UC#3 - Facilitating urban policy making and monitoring through crowdsourcing data analysis	76
3.1.4	UC#4 - Predictive analysis towards unemployment risks identification and policy making..	77
4	Platform Roles.....	79
5	System Requirements	80
5.1	Cloud Capabilities and Data Collection	83
5.2	Data Governance Model and Incentives Management.....	84
5.3	Reusable Models and Analytical Tools	90
5.4	Policy Management Framework.....	96
5.5	Policy Development Toolkit	98
5.6	Data Marketplace	103
5.7	Legal and Ethical Framework.....	107
6	Software Requirements.....	110
6.1	Cloud Provisioning	110
6.2	Cloud Register	111
6.3	Cloud Gateways	112
6.4	Incentives Management	119
6.5	Data Cleaning.....	120
6.6	Data Fusion Linking.....	131
6.7	Data Interoperability	134
6.8	Data Store.....	138
6.9	Opinion Mining	142
6.10	Sentiment Analysis	144
6.11	Behavioural Analysis	147
6.12	Situational Knowledge Analysis	150
6.13	Optimization and Reusability.....	153
6.14	Policy Development Toolkit	159
6.15	Data Visualization	164
7	State of the Art analysis.....	165
7.1	Evidence based policy making and data analytics	165
7.2	Policy interoperable datasets	169
7.3	Enhanced visualizations providing actionable insights.....	170

7.4	Re-usability of analytical tools/models decoupled from infrastructure	173
7.5	Polyglot analytical tools federating heterogeneous sources and stores.....	174
7.6	Efficient data fusion from various data sources	175
7.7	Efficient cloud infrastructures	177
7.8	Data Governance Model, Protection and Privacy Enforcement	178
7.9	Cross-sector Policy Lifecycle Management.....	180
8	Background Technologies	183
9	Conclusion	186
	References	187

List of Tables

Table 1 - WORK PRODUCTS DESCRIPTION	19
Table 2: SCENARIO SCE-PPR-01	23
Table 3 - SCENARIO SCE-PPR-02.....	23
Table 4 - SCENARIO SCE-PPR-03.....	24
Table 5 - SCENARIO SCE-PPR-04.....	25
Table 6 - SCENARIO SCE-PPR-06.....	26
Table 7 - SCENARIO SCE-PPR-08.....	26
Table 8 - STAKEHOLDER REQUIREMENT REQ-PPR-01	27
Table 9 - STAKEHOLDER REQUIREMENT REQ-PPR-02	27
Table 10 - STAKEHOLDER REQUIREMENT REQ-PPR-03	28
Table 11 - STAKEHOLDER REQUIREMENT REQ-PPR-04.....	28
Table 12 - STAKEHOLDER REQUIREMENT REQ- PPR-05.....	29
Table 13 - STAKEHOLDER REQUIREMENT REQ- PPR-06.....	29
Table 14 - STAKEHOLDER REQUIREMENT REQ-PPR-07	30
Table 15 - STAKEHOLDER REQUIREMENT REQ-PPR-08.....	30
Table 16 - STAKEHOLDER REQUIREMENT REQ-PPR-09.....	30
Table 17: STAKEHOLDER REQUIREMENT REQ-PPR-10	31
Table 18 - STAKEHOLDER REQUIREMENT REQ-PPR-11	31
Table 19 - STAKEHOLDER REQUIREMENT REQ-PPR-12.....	32
Table 20 - STAKEHOLDER REQUIREMENT REQ-PPR-13.....	32
Table 21 - STAKEHOLDER REQUIREMENT REQ-PPR-14.....	33
Table 22 - STAKEHOLDER REQUIREMENT REQ-PPR-15.....	33
Table 23 - SCENARIO SCE-IIPDD-01.....	35
Table 24 - SCENARIO SCE-IIPDD-02.....	36
Table 25 - SCENARIO SCE-IIPDD-03.....	37
Table 26 - SCENARIO SCE-IIPDD-04.....	38
Table 27 - SCENARIO SCE-IIPDD-05.....	39
Table 28 - SCENARIO SCE-IIPDD-06.....	40
Table 29 - SCENARIO SCE-IIPDD-07.....	41
Table 30 - SCENARIO SCE-IIPDD-08.....	42
Table 31 - SCENARIO SCE-IIPDD-09.....	43
Table 32 - SCENARIO SCE-IIPDD-10.....	44
Table 33 - SCENARIO SCE-IIPDD-11.....	45
Table 34 - STAKEHOLDER REQUIREMENT REQ- IIPDD-01	45
Table 35 - STAKEHOLDER REQUIREMENT REQ- IIPDD-02.....	46
Table 36 - STAKEHOLDER REQUIREMENT REQ- IIPDD-03.....	46
Table 37 - STAKEHOLDER REQUIREMENT REQ- IIPDD-05.....	47
Table 38 - STAKEHOLDER REQUIREMENT REQ- IIPDD-06.....	48
Table 39 - STAKEHOLDER REQUIREMENT REQ- IIPDD-07.....	48
Table 40 - STAKEHOLDER REQUIREMENT REQ- IIPDD-08.....	49

Table 41 - STAKEHOLDER REQUIREMENT REQ- IIPDD-09.....	49
Table 42 - STAKEHOLDER REQUIREMENT REQ- IIPDD-11.....	49
Table 43 - STAKEHOLDER REQUIREMENT REQ- IIPDD-12.....	50
Table 44 - STAKEHOLDER REQUIREMENT REQ- IIPDD-13.....	50
Table 45 - STAKEHOLDER REQUIREMENT REQ- IIPDD-14.....	51
Table 46 - STAKEHOLDER REQUIREMENT REQ- IIPDD-15.....	51
Table 47 - STAKEHOLDER REQUIREMENT REQ- IIPDD-16.....	52
Table 48 - STAKEHOLDER REQUIREMENT REQ- IIPDD-17.....	52
Table 49 - SCENARIO SCE-RLIMP-01	55
Table 50 - SCENARIO SCE-RLIMP-02	56
Table 51 - SCENARIO SCE-RLIMP-03	57
Table 52 - SCENARIO SCE-RLIMP-04	58
Table 53 - SCENARIO SCE-RLIMP-05	59
Table 54 - SCENARIO SCE-RLIMP-06	60
Table 55 - STAKEHOLDER REQUIREMENT REQ- RLIMP-01	60
Table 56 - Scenario SCE-PAUNRI-01	62
Table 57 - Scenario SCE-PAUNRI-01	63
Table 58 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-01	64
Table 59 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-02	65
Table 60 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-03	65
Table 61 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-04	66
Table 62 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-05	66
Table 63 - DATASET REQUIREMENT DS -PPR-01	67
Table 64 - DATASET REQUIREMENT DS-PPR-02	68
Table 65 - DATASET REQUIREMENT FOR DS-PPR-03	68
Table 66 - DATASET REQUIREMENT FOR DS-PPR-04	69
Table 67 - DATASET REQUIREMENT FOR DS-PPR-05	70
Table 68 - DATASET REQUIREMENT FOR DS-IIPDD-01	71
Table 69 - DATASET REQUIREMENT FOR DS-IIPDD-02	72
Table 70 - DATASET REQUIREMENT FOR DS-IIPDD-03	72
Table 71 - DATASET REQUIREMENT FOR DS-IIPDD-04	73
Table 72 - DATASET REQUIREMENT FOR DS-IIPDD-05	73
Table 73 - DATASET REQUIREMENT FOR DS-IIPDD-06	74
Table 74 - DATASET REQUIREMENT FOR DS-IIPDD-08	74
Table 75 - DATASET REQUIREMENT FOR DS-IIPDD-09	75
Table 76 - DATASET REQUIREMENT FOR DS-IIPDD-10	75
Table 77 - DATASET REQUIREMENT FOR DS-RLIMP-01	76
Table 78 - DATASET REQUIREMENT FOR DS-RLIMP-02	77
Table 79 - DATASET REQUIREMENT FOR DS-PAUNRI-01	78
Table 80 - PLATFORM ROLES	79
Table 81 - SYSTEM REQUIREMENT REQ -SY-CCDC-01	83

Table 82 - SYSTEM REQUIREMENT REQ-SY-CCDC-02	83
Table 83 - SYSTEM REQUIREMENT REQ - SY-DGMIM-01	84
Table 84 - SYSTEM REQUIREMENT REQ- SY-DGMIM-02	85
Table 85 - SYSTEM REQUIREMENT REQ- SY-DGMIM-03	86
Table 86 - SYSTEM REQUIREMENT REQ- SY-DGMIM-05	87
Table 87 - SYSTEM REQUIREMENT REQ- SY-DGMIM-06	87
Table 88 - SYSTEM REQUIREMENT REQ- SY-DGMIM-07	88
Table 89 - SYSTEM REQUIREMENT REQ- SY-DGMIM-08	88
Table 90 - SYSTEM REQUIREMENT REQ- SY-DGMIM-09	89
Table 91 - SYSTEM REQUIREMENT REQ- SY-DGMIM-10	89
Table 92 - SYSTEM REQUIREMENT REQ- SY-DGMIM-11	90
Table 93 - SYSTEM REQUIREMENT REQ -SY-RMAT-01	91
Table 94 - SYSTEM REQUIREMENT REQ -SY-RMAT-02	91
Table 95 - SYSTEM REQUIEREMENT REQ-SY-RMAT-04	92
Table 96 - SYSTEM REQUIEREMENT REQ-SY-RMAT-05	93
Table 97 - SYSTEM REQUIEREMENT REQ-SY-RMAT-06	93
Table 98 - SYSTEM REQUIEREMENT REQ-SY-RMAT-07	93
Table 99 - SYSTEM REQUIEREMENT REQ-SY-RMAT-08	94
Table 100 - SYSTEM REQUIEREMENT REQ-SY-RMAT-09	94
Table 101 - SYSTEM REQUIEREMENT REQ-SY-RMAT-10	95
Table 102 - SYSTEM REQUIREMENT REQ-SY-PMF-01	96
Table 103 - SYSTEM REQUIREMENT REQ-SY-PMF-02	97
Table 104 - SYSTEM REQUIREMENT REQ-SY-PMF-03	97
Table 105 - SYSTEM REQUIREMENT REQ-SY-PMF -04	97
Table 106 - SYSTEM REQUIREMENT REQ - SY-PDT-01	98
Table 107 - SYSTEM REQUIREMENT REQ - SY-PDT-02	98
Table 108 - SYSTEM REQUIREMENT REQ- SY-PDT-03	99
Table 109 - SYSTEM REQUIREMENT REQ- SY-PDT-05	99
Table 110 - SYSTEM REQUIREMENT REQ- SY-PDT-06	100
Table 111 - SYSTEM REQUIREMENT REQ- SY-PDT-07	100
Table 112 - SYSTEM REQUIREMENT REQ- SY-PDT-08	101
Table 113 - SYSTEM REQUIREMENT REQ- SY-PDT-09	101
Table 114 - SYSTEM REQUIREMENT REQ- SY-PDT-10	102
Table 115 - SYSTEM REQUIREMENT REQ- SY-PDT-12	103
Table 116 - SYSTEM REQUIREMENT REQ- SY-DM-01	104
Table 117 - SYSTEM REQUIREMENT REQ- SY-DM-02	104
Table 118 - SYSTEM REQUIREMENT REQ- SY-DM-03	104
Table 119 - SYSTEM REQUIREMENT REQ- SY-DM-04	105
Table 120 - SYSTEM REQUIREMENT REQ- SY-DM-05	106
Table 121 - SYSTEM REQUIREMENT REQ- SY-DM-06	106
Table 122 - SYSTEM REQUIREMENT REQ- SY-DM-07	106

Table 123 - SYSTEM REQUIREMENT REQ- SY-DM-08	107
Table 124 - System Requirement REQ- SY-LEF-01	109
Table 125 - SOFTWARE REQUIREMENT REQ - SO-CP-01	111
Table 126 - SOFTWARE REQUIREMENT REQ - SO-CR-01	111
Table 127 - SOFTWARE REQUIREMENT REQ -SO-CG-01.....	112
Table 128 - SOFTWARE REQUIREMENT REQ -SO-CG-02.....	112
Table 129 - SOFTWARE REQUIREMENT REQ -SO-CG-03.....	113
Table 130 - SOFTWARE REQUIREMENT REQ -SO-CG-04.....	113
Table 131 - SOFTWARE REQUIREMENT REQ -SO-CG-05.....	114
Table 132 - SOFTWARE REQUIREMENT REQ -SO-CG-06.....	114
Table 133 - SOFTWARE REQUIREMENT REQ -SO-CG-07.....	115
Table 134 - SOFTWARE REQUIREMENT REQ -SO-CG-08.....	115
Table 135 - SOFTWARE REQUIREMENT REQ -SO-CG-09.....	116
Table 136 - SOFTWARE REQUIREMENT REQ -SO-CG-10.....	116
Table 137 - SOFTWARE REQUIREMENT REQ -SO-CG-11.....	117
Table 138 - SOFTWARE REQUIREMENT REQ -SO-CG-12.....	117
Table 139 - SOFTWARE REQUIREMENT REQ -SO-CG-13.....	118
Table 140 - SOFTWARE REQUIREMENT REQ -SO-CG-14.....	118
Table 141 - SOFTWARE REQUIREMENT REQ -SO-CG-15.....	119
Table 142 - SOFTWARE REQUIREMENT REQ -SO-CG-16.....	119
Table 143 - SOFTWARE REQUIREMENT REQ - SO-IM-01	120
Table 144 - SOFTWARE REQUIREMENT REQ - SO-IM-02	120
Table 145 - SOFTWARE REQUIREMENT REQ- SO-DC-01	121
Table 146 - SOFTWARE REQUIREMENT REQ- SO-DC-02.....	121
Table 147 - SOFTWARE REQUIREMENT REQ- SO-DC-03.....	122
Table 148 - SOFTWARE REQUIREMENT REQ- SO-DC-04.....	122
Table 149 - SOFTWARE REQUIREMENT REQ- SO-DC-05.....	123
Table 150 - SOFTWARE REQUIREMENT REQ- SO-DC-06.....	123
Table 151 - SOFTWARE REQUIREMENT REQ- SO-DC-07.....	124
Table 152 - SOFTWARE REQUIREMENT REQ- SO-DC-08.....	124
Table 153 - SOFTWARE REQUIREMENT REQ- SO-DC-09.....	125
Table 154 - SOFTWARE REQUIREMENT REQ- SO-DC-10.....	125
Table 155 - SOFTWARE REQUIREMENT REQ- SO-DC-11.....	126
Table 156 - SOFTWARE REQUIREMENT REQ- SO-DC-12.....	126
Table 157 - SOFTWARE REQUIREMENT REQ- SO-DC-13.....	127
Table 158 - SOFTWARE REQUIREMENT REQ- SO-DC-14.....	127
Table 159 - SOFTWARE REQUIREMENT REQ- SO-DC-15.....	128
Table 160 - SOFTWARE REQUIREMENT REQ- SO-DC-16.....	128
Table 161 - SOFTWARE REQUIREMENT REQ- SO-DC-17.....	129
Table 162 - SOFTWARE REQUIREMENT REQ- SO-DC-18.....	129
Table 163 - SOFTWARE REQUIREMENT REQ- SO-DC-20.....	130

Table 164 - SOFTWARE REQUIREMENT REQ- SO-DC-21	130
Table 165 - SOFTWARE REQUIREMENT REQ - SO-DFL-01	131
Table 166 - SOFTWARE REQUIREMENT REQ - SO-DFL-02	131
Table 167 - SOFTWARE REQUIREMENT REQ - SO-DFL-03	132
Table 168 - SOFTWARE REQUIREMENT REQ - SO-DFL-04	132
Table 169 - SOFTWARE REQUIREMENT REQ - SO-DFL-05	133
Table 170 - SOFTWARE REQUIREMENT REQ - SO-DFL-06	133
Table 171 - SOFTWARE REQUIREMENT REQ - SO-DI-01	134
Table 172 - SOFTWARE REQUIREMENT REQ - SO-DI-02	134
Table 173 - SOFTWARE REQUIREMENT REQ - SO-DI-03	135
Table 174 - SOFTWARE REQUIREMENT REQ - SO-DI-04	135
Table 175 - SOFTWARE REQUIREMENT REQ - SO-DI-05	136
Table 176 - SOFTWARE REQUIREMENT REQ - SO-DI-06	136
Table 177 - SOFTWARE REQUIREMENT REQ - SO-DI-07	137
Table 178 - SOFTWARE REQUIREMENT REQ - SO-DI-08	137
Table 179 - SOFTWARE REQUIREMENT REQ - SO-DI-09	138
Table 180 - SOFTWARE REQUIREMENT REQ - SO-DS-01	138
Table 181 - SOFTWARE REQUIREMENT REQ - SO-DS-02	139
Table 182 - SOFTWARE REQUIREMENT REQ - SO-DS-03	140
Table 183 - SOFTWARE REQUIREMENT REQ - SO-DS-05	140
Table 184 - SOFTWARE REQUIREMENT REQ - SO-DS-06	141
Table 185 - SOFTWARE REQUIREMENT REQ - SO-DS-07	142
Table 186 - SOFTWARE REQUIREMENT REQ - SO-OM-01	142
Table 187 - SOFTWARE REQUIREMENT REQ - SO-OM-03	143
Table 188 - SOFTWARE REQUIREMENT REQ - SO-OM-04	143
Table 189 - SOFTWARE REQUIREMENT REQ - SO-OM-07	144
Table 190 - SOFTWARE REQUIREMENT REQ - SO-OM-08	144
Table 191 - SOFTWARE REQUIREMENT REQ - SO-OM-09	144
Table 192 - SOFTWARE REQUIREMENT REQ - SO-SA-01	145
Table 193 - SOFTWARE REQUIREMENT REQ - SO-SA-02	145
Table 194 - SOFTWARE REQUIREMENT REQ - SO-SA-03	146
Table 195 - SOFTWARE REQUIREMENT REQ - SO-SA-04	146
Table 196 - SOFTWARE REQUIREMENT REQ - SO-SA-05	147
Table 197 - SOFTWARE REQUIREMENT REQ - SO-BA-01	147
Table 198 - SOFTWARE REQUIREMENT REQ - SO-BA-02	148
Table 199 - SOFTWARE REQUIREMENT REQ - SO-BA-03	148
Table 200 - SOFTWARE REQUIREMENT REQ - SO-BA-04	149
Table 201 - SOFTWARE REQUIREMENT REQ - SO-BA-05	149
Table 202 - SOFTWARE REQUIREMENT REQ - SO-SKA-01	150
Table 203 - SOFTWARE REQUIREMENT REQ - SO-SKA-02	150
Table 204 - SOFTWARE REQUIREMENT REQ - SO-SKA-03	151

Table 205 - SOFTWARE REQUIREMENT REQ - SO-SKA-04	151
Table 206 - SOFTWARE REQUIREMENT REQ - SO-SKA-05	152
Table 207 - SOFTWARE REQUIREMENT REQ - SO-SKA-06	152
Table 208 - SOFTWARE REQUIREMENT REQ - SO-OR-01	153
Table 209 - SOFTWARE REQUIREMENT REQ - SO-OR-02	154
Table 210 - SOFTWARE REQUIREMENT REQ - SO-OR-03	154
Table 211 - SOFTWARE REQUIREMENT REQ- SO-OR-04	155
Table 212 - SOFTWARE REQUIREMENT REQ- SO-OR-05	156
Table 213 - SOFTWARE REQUIREMENT REQ- SO-OR-06	156
Table 214 - SOFTWARE REQUIREMENT REQ- SO-OR-07	157
Table 215 - SOFTWARE REQUIREMENT REQ- SO-OR-08	158
Table 216 - SOFTWARE REQUIREMENT REQ- SO-OR-09	158
Table 217 - SOFTWARE REQUIREMENT REQ- SO-OR-10	159
Table 218 - Software Requirement REQ-SO-PDT-01	159
Table 219 - SOFTWARE REQUIREMENT REQ-SO-PDT-02	160
Table 220 - SOFTWARE REQUIREMENT REQ-SO-PDT-03	160
Table 221 - SOFTWARE REQUIREMENT REQ-SO-PDT-04	161
Table 222 - SOFTWARE REQUIREMENT REQ-SO-PDT-05	161
Table 223 - SOFTWARE REQUIREMENT REQ-SO-PDT-06	161
Table 224 - SOFTWARE REQUIREMENT REQ-SO-PDT-07	162
Table 225 - SOFTWARE REQUIREMENT REQ-SO-PDT-08	162
Table 226 - SOFTWARE REQUIREMENT REQ-SO-PDT-09	163
Table 227 - SOFTWARE REQUIREMENT REQ-SO-PDT-10	163
Table 228 - SOFTWARE REQUIREMENT REQ-SO-PDT-11	164
Table 229 - Software Requirement REQ-SO-DV-01	164
Table 230 - MOST POPULAR JS CHART LIBRARIES [19], [20], [21], [22], [23]	172
Table 231 - MOST POPULAR JS FRAMEWORKS [24], [25]	172
Table 232 - COMPATIBILITY BETWEEN JS LIBRARIES AND FRAMEWORKS	173
Table 233 - BASELINE TECHNOLOGIES	184
Table 234 - RELEVANT RESEARCH PROJECTS	185

List of Figures

Figure 1 - METHODOLOGY FOR REQUIREMENTS ENGINEERING	18
Figure 2 - POLICYCLOUD CORE SYSTEM CAPABILITIES	80
Figure 3: POLICY MAKING CYCLE	165
Figure 4 - RESEARCH CLUSTERS ACCORDING TO THE BIG POLICY CANVAS PROJECT	167

Executive Summary

This deliverable summarizes the work that has been done in the scope of T2.1 (“Requirements Elicitation & State of the Art Analysis”). It is the third and last version of the series of deliverables of this task, whose main objective is to specify the use case scenarios, their involved datasets and their relevant user requirements, as well as the system and technical requirements that are being imposed by the platform. The purpose of this series is to track those requirements throughout the project and update them during the progress of the project. As a result, this third and last version depicts the final technical requirements that have been identified after the delivery of the first two versions of the scientific reports, providing an extended version of the user scenarios, the definition of their datasets, along with an extensive survey regarding the regulatory constraints. As mentioned in the previous versions of this deliverable, the approach that has been followed is twofold: A top-down approach has been followed with respect to the user requirements that were initially collected by the use case providers themselves, after specifying the business goals and objectives of the use case, along with a concrete definition of the scenario. In this version, additional requirements have been identified, after the successful demonstration of the initial prototype that took place during this period (January 2021 – June 2021), being also demonstrated in the first review of the project, gathering additional information coming from the engagement of end users of the domain that provided valuable input, which is reflected in this last version of the deliverable. Moreover, a bottom-up approach has been applied, whose objective is to additionally complement and identify/analyse the technical requirements, which have been updated with respect to the progress that has been made by the work carried out in the technical work packages, which are focusing on the platform technological needs.

The result of this analysis is the updated list of measurable unambiguous requirements that validates the initial design of the overall architecture of the PolicyCLOUD platform, focusing on serving all the different needs of the various use cases of the project, with the target to be aligned with the requirements defined in this document. As the project is still progressing, this final version of this deliverable will help the architecture designers of the platform and its software developers to adjust and validate the overall architecture and its implementation accordingly, as the principles of the agile methodology demand. Moreover, as our scope is to continuously keep track with the latest technological advances and how we can take advantage of them while developing our platform, the analysis of the state-of-the-art analysis regarding the major technologies that are envisioned to be exploited has been extended, while this document maintains the list of several projects whose technological assets might be candidates to be incorporated in the overall solution.

1 Introduction

This document purpose is to provide an updated list of the measurable and specific **user, business and system requirements** that drives the design of the architecture of the PolicyCLOUD platform and has been used as the basis for the validation and further implementation of the relevant functionalities that will be offered by the various software components of the platform.

This report is the third and final version of this series of deliverables and reports the work that is being carried out in PolicyCLOUD's task T2.1 "*Requirements Elicitation & State of the Art Analysis*", whose main goal is to collect the user and system requirements and track them during the course of the project. The analysis of the requirements is producing **a measurable and unambiguous requirement set**, which is currently being tracked against the developments during the project lifecycle in order to ensure that the PolicyCLOUD complexity will be fully addressed and properly considered. Moreover, another important goal of task T2.1 is **to investigate and analyse the State-of-the-Art (SotA)** for PolicyCLOUD technologies. Both these two goals provide valuable input for the design of the overall platform architecture and all the research activities of the project. As the importance of this task is high, since it affects the overall design of the project, the task started in January 2020 (M1). An internal version of that report was released in March 2020 (M3) and was further refined during the first period. D2.1 State of the Art & Requirements Analysis, delivered in June 2020, reflected what had been done until M6. Then, a second version (D2.4 State of the Art & Requirements Analysis M12) of the document was released in December 2020, containing an updated list of requirements both from a user and from a technology perspective, along with updates of the state-of-the-art technologies that could be exploited in the project. In this last version of the document, an updated list of requirements has been provided, along with further extensions and modifications of the user scenarios, their involved datasets and their regulatory constraints.

The analysis and elicitation of the requirements have been carried out taking into account the exact needs and concerns that have been identified by the current communities, end-users and related actors that are related to the PolicyCLOUD use cases and providers of the corresponding technologies. This was possible through dedicated virtual meetings with the users organized for each pilot, due to the COVID-19 restrictions in most European countries, it was difficult to organize big events as it was planned. As a result, the analysis that has been made not only specifies **use case requirements**, which can also be considered as stakeholder requirements under ISO/IEC/IEEE 29148:2011 [1], but also **technical requirements** that can be considered as system and software requirements. At the time that this version of the deliverable is published, October 2021 (M22), the system requirements have been extended as results of the technical work that has been carried out in the technical work packages of the project. As during the first phase of the project, most of the use cases related activities have to deal with the unprecedented task of managing the pandemic crisis, which was at its peak of the first wave in Europe, it was very difficult to collect the description of the use cases along with the additional information regarding the types of different datasets that will be brought to the platform. Due to this, in this version of the deliverable, an extended analysis of the use cases has been provided. Moreover, the main focus

has also been given on the updates of the definition of the software technology requirements that are being imposed by each of the main components. Due to the complexity of the overall architecture and the wide variety of the technologies used in the PolicyCLOUD platform, the system requirements collected from users and the software ones, have been further developed in this final version of this deliverable.

Apart from the list of the functional and non-functional requirements, this deliverable was planned to have an additional section to describe the **various categories of the stakeholders of the PolicyCLOUD platform, the different envisioned business models and the expected business outcomes and business goals**. This analysis, even if it is not addressing strictly technical perspectives of the project, it was considered to be a valuable input for the activities of ‘WP7 Communication, Exploitation, Standardisation, Roadmapping & Business Development’, mainly concerning the project’s road mapping and business development. However, the results of this analysis have not been included in this document since they are more aligned with WP7 activities and results. Moreover, to better understand the software technology requirements, this deliverable includes an analysis of the related technologies state-of-the-art. At the same time, an initial and non-exhaustive list of relevant research initiatives and projects is provided, along with the description of the baseline technologies that the technical partners will bring to the project.

This document is organized as follows: Section 1.1 explains the method that has been followed to gather the platform requirements, while Section 1.2 states the differences and updates with the previous versions of this deliverable; Section 2 provides a brief description of the scenarios of the different use cases, along with the list of the user requirements, while Section 3 describes the various datasets used in the context of PolicyCLOUD. Section 4 defines the different roles of the users of the PolicyCLOUD ecosystem, while Sections 5 and 6 provide the technical requirements of the platform. Sections 7 and 8 present the state-of-the-art analysis and specify a list of the baseline technologies used in the development and implementation phases of the platform, while Section 9 finally concludes the document.

1.1 Method

The engineering method to gather all the user and technical requirements for the PolicyCLOUD project follows the ISO/IEC/IEEE 29148:2011 norm, as already mentioned in the previous subsection. This norm describes two main processes and practices that need to be executed in an iterative and recursive manner.

The first process is related to the **definition of the requirements coming from the stakeholders**. Its purpose is to define the requirements for a system that can provide the services needed by users and other stakeholders in a defined environment. The output of this process is the *Stakeholder Requirements Specification (StRS)*. On the other hand, the second process that is defined in the norm is related with the **requirements analysis** whose main purpose is to transform the stakeholder, requirement-driven view of desired services into a technical view of a required product that could deliver those services. The outcome

of the second process is *the System Requirements Specification (SyRS)* and the *Software Requirements Specification (SRS)*.

The specification of the overall requirement can be provided in three levels of detail, which serve as input to different practices or stages in the architectural design process. Their relations can be seen in Figure 1. Moreover, with respect to the ISO/IEC/IEEE 29148:2011, Table 1 describes the relations of each of the three outcomes of the two processes that are defined - the Stakeholder Requirements Specification (StRS), the System Requirements Specification (SyRS) and the Software Requirements Specification (SRS), including the architecture domain whose decisions are informed by them.

PolicyCLOUD Requirements Methodology

Iterative and recursive method

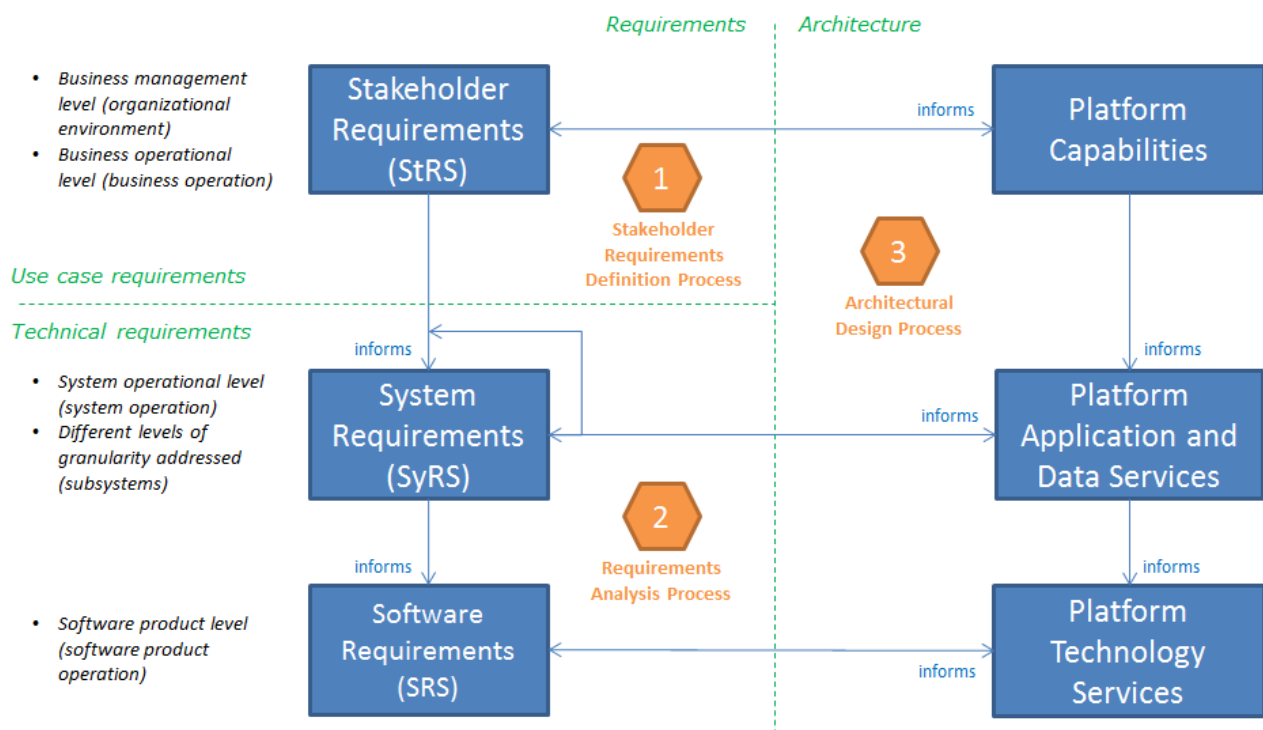


FIGURE 1 - METHODOLOGY FOR REQUIREMENTS ENGINEERING

Work Product	Acronym	Description	Informed Domain	Architecture
Business Requirements Specification	StRS	This contains the requirements as defined by the use case providers	Platform Capabilities (business architecture)	
System Requirements Specification	SyRS	This defines the platform level requirements	Platform Applications and Data Services Architecture	
Software Requirements Specification	SRS	This contains the specific requirements of each one of the individual components	Platform Technology Architecture	

TABLE 1 - WORK PRODUCTS DESCRIPTION

Moreover, to highlight the key business requirements and indicate the implied technical requirements for the overall architecture of the PolicyCLOUD platform, we used the *TOGAF® Series Guide - Business Scenarios*¹ methodology that facilitates the identification of the requirements from the stakeholders' point of view. This technique aims to validate, elaborate and modify the premise behind an architecture effort, by focusing on the understanding and documentation of the key elements of a business scenario using iterations in an agile manner.

Finally, to better formalize the requirements, we use the following attributes:

- **Level of detail:** Following the use of ISO/IEC/IEEE 29148:2011, we use the following levels - Stakeholder, System and Software (i.e., technology details).
- **Type:** Types of requirements are functional - FUNC (function), DATA (data), non-functional - L&F (Look and Feel Requirements), USE (Usability Requirements), PERF (Performance Requirements), ENV (Operational/Environment Requirements), and SUP (Maintainability and Support Requirements).
- **Priority:** Requirements can have different priorities - MAN (mandatory requirement), DES (desirable requirement), OPT (optional requirement), ENH (possible future enhancement).

¹ <https://publications.opengroup.org/g176>

1.2 Summary of Changes

This section highlights the updates made to the previous version of this deliverable, D2.4 State of the Art & Requirements Analysis:

- Section 2:
 - New scenarios defined for LON and SOFIA: SCE-RLIMP-06, SCE-PAUNRI-01, SCE-PAUNRI-02
 - Updates and refinements in the following scenarios: SCE-IIPDD-02, SCE-IIPDD-05, SCE-PPR-01, SCE-PPR-02, SCE-PPR-03, SCE-PPR-04, SCE-PPR-06, SCE-PPR-08
 - UML diagrams provided in all scenarios
 - The following scenarios are not valid: SCE-PPR-05, SCE-PPR-07
 - Updates on requirements: all from REQ- PPR-01 to REQ- PPR-015, all for ITA/SARGA, REQ-PAUNRI-01, REQ-PAUNRI-02, REQ-PAUNRI-03, REQ-PAUNRI-04
 - New requirement added: REQ-PAUNRI-05
- Section 3:
 - Updates on the description of the datasets: DS-PPR-01, DS-PPR-02, DS-IIPDD-03, DS-RLIMP-02
 - Dataset DS-IIPDD-07 is not valid
- Section 5:
 - New section regarding the marketplace and the *policy management framework*
 - New requirements: REQ- SY-PDT-12, REQ- SY-PDT-13, REQ- SY-PDT-14, REQ-SY-DM-01, REQ-SY-DM-02, REQ-SY-DM-03, REQ-SY-DM-04, REQ-SY-DM-05, REQ-SY-DM-06, REQ-SY-DM-07, REQ-SY-DM-08, REQ-SY-PMF-01, REQ-SY-PMF-0, REQ-SY-PMF-03, REQ-SY-PMF-04
 - Updates on requirements: REQ- SY-CCDC-02, REQ- SY-RMAT-04
 - Removal of requirements: REQ-SY-PDT-04, REQ-SY-PDT-11, REQ-SY-PDT-13, REQ-SY-PDT-14
 - Update on the description about the Legal and Ethical Framework
- Section 6:
 - New requirements: REQ- SO-CG-12, REQ- SO-CG-13, REQ- SO-CG-14, REQ- SO-CG-15, REQ- SO-CG-16, REQ-SO-DS-07, REQ- SO-OM-07, REQ- SO-OM-08, REQ- SO-OM-09, REQ- SO-SA-03, REQ- SO-SA-04, REQ- SO-SA-05, REQ- SO-SKA-04, REQ- SO-SKA-05, REQ- SO-SKA-06, REQ- SO-OR-10, REQ- SO-PM-02, REQ- SO-PM-03, REQ- SO-PM-04, REQ-SO-PDT-09, REQ-SO-PDT-10, REQ-SO-PDT-11
 - Updates on requirements: REQ-SO-SKA-01, REQ-SO-BA-04, REQ-SO-BA-05, REQ-SO-SA-01, REQ-SO-SA-02, REQ- SO-OM-03, REQ- SO-OM-04, REQ-SO-DS-05, REQ-SO-DS-06, REQ- SO-DC-01 to REQ- SO-DC-18 and REQ- SO-DC-20 to REQ- SO-DC-21
 - Removal of requirements: REQ- SO-DC-19, REQ- SO-OM-02, REQ- SO-OM-05, REQ- SO-OM-06
 - Old section 6.14 “Policy Modeling” has been removed and merged with the “Policy Development Toolkit”.

2 Use case Requirements and Scenarios

The purpose of this section is to present the business usage scenarios along with the list of the initial requirements that have been defined by each of the four use cases of the PolicyCLOUD project. These requirements formulate the list of the overall Stakeholder Requirements, according to the engineering methodology that has been analysed in Table 1.

Each of the four cases describes the exact usage from a use case perspective at a high-level description. It is worth to be mentioned that the complete definition of the detailed scenario that is needed was not the focus of this analysis, as this is part of the work will be carried out in the scope of T6.2 *“Use case Definition & Design”*. The scope of the work that is being reported in this section is rather the general descriptions that are more related with the general definition of the behaviour and identification of the important necessities that the architecture should comply with, so that they can be taken into account from the very beginning of the project. The descriptions of the scenarios are complemented with UML Use Case Diagrams in order to identify the different actors, prerequisites and descriptions of the behaviour.

The following subsections firstly provide an introductory overview of the purpose of each scenario, followed by a more detailed description of the use case. Then, the description of the different user stories that formulate each scenario is presented, along with the corresponding UML diagram, and finally, the initial list of the stakeholder requirements is reported.

2.1 UC#1: Participatory policies against radicalization

2.1.1 Goals and Objectives

UC#1 aims to develop a collaborative data-driven application for the evaluation of public policies to counter and prevent radicalization based on a participatory review of data coming from social media (Twitter, Facebook, Reddit) and open datasets (GDT, RDWTI). In addition, it will provide useful insights to policy makers at local, regional and national level to update existing policies and investigate whether new ones are needed.

This use case will address the challenge of radicalisation by offering policy makers ICT-based tools for enabling them to **identify, monitor, analyse, and predict potential risks of radicalization**, while at the same time **allow them to interact with other stakeholders** (i.e., data analytics professionals, social scientists, legal experts) during the creation and modelling of policies and measures to counter violent extremism.

UC#1 will adopt the PolicyCLOUD technologies developed during the project, and in particular:

- The **PolicyCLOUD opinion mining and sentiment analysis** tools, which provide:
 - social media analysis towards the identification of radicalization activities and involved actors (individual, groups), as well as linking of data about terroristic groups and attacks with radicalization efforts.

- Alleviate the negative consequences of counter-radicalisation policies (e.g., restrictions) by making them more targeted.
 - big data analytics to identify origins of radicalization efforts (including countries/regions and terroristic groups conducting them), risk assignment probabilities to suspects of radicalization efforts, and segmentation of radicalization efforts and subjects based on demographics and risks.
 - opinion/sentiment classification and user type classification, as well as pattern identification and analysis over time will also be enabled through approaches developed in this task.
- The **PolicyCLOUD situational knowledge acquisition and analysis** tools, which provide:
 - knowledge acquisition from real-world data using statistical algorithms and machine learning techniques in combination with collective knowledge (out of the clusters/collections of policies – Task 5.5), and predictive risk analysis. The knowledge derived will influence different types of proposed decisions towards target communities given the developed knowledge regarding efficiency of decisions, current status and policy planning.
- The **PolicyCLOUD visualization** technologies, which will:
 - enable policy makers to depict visually the radicalisation trends and poles.

2.1.2 Description of Scenarios

The following tables define the various scenarios of this use case.

Section	Description
ID	SCE-PPR-01
Title	Social Media Configuration
Description	Insert new keywords to search
Actors	Administrator, Policy makers
Objectives	Describing elements from social networks where information has to be analysed
Pre-Conditions	<ul style="list-style-type: none"> • The Administrator should have been identified on the platform • Social networks used: Twitter, Facebook, Reddit, RSS Feed
Process Description	<ul style="list-style-type: none"> • Administrator selects “My research interests” • System shows a wizard where the user enters configuration parameters - social networks, topics to inspect • Administrator inserts, modifies or deletes social networks and keywords • System saves configuration and adds new parameters to be used by the probe
Variations	N/A
Post-Conditions	<ul style="list-style-type: none"> • Text is identified and analysed • Hierarchical text categorization defined (taxonomy or ontology)

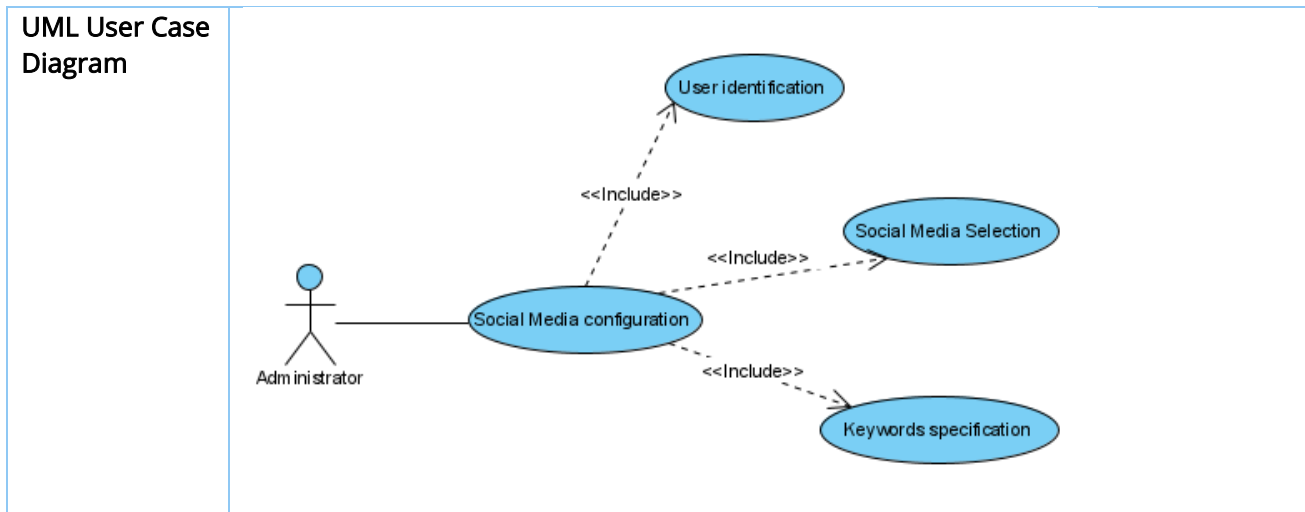


TABLE 2: SCENARIO SCE-PPR-01

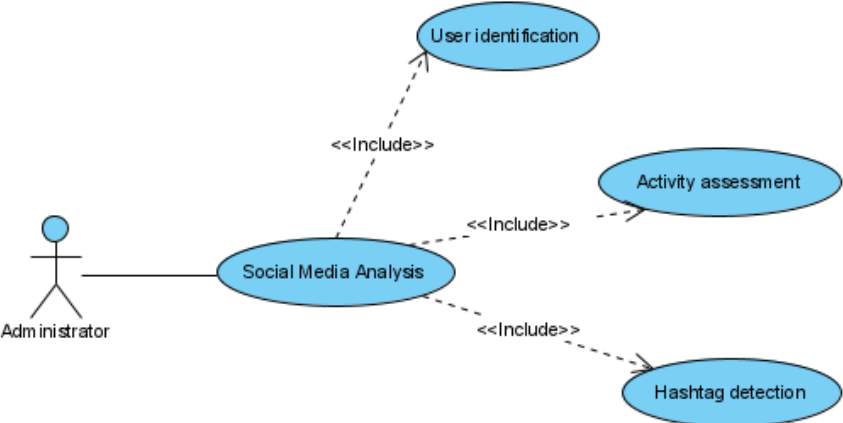
Section	Description
ID	SCE-PPR-02
Title	Social Media Analysis
Description	Identify the actors (individual, groups) involved in radicalization efforts
Actors	Analyst, Policy maker
Objectives	<ul style="list-style-type: none"> Identify users and groups on several social networks Identify activities on posts (comments, likes, replies, etc.) New keywords identification
Pre-Conditions	Verify activity on platforms - Twitter, Reddit, RSS Feed
Process Description	<ul style="list-style-type: none"> System read the text to process System analysed and categorized the text
Variations	N/A
Post-Conditions	<ul style="list-style-type: none"> Text is identified and analysed Text categorization defined (taxonomy or ontology)
UML User Case Diagram	 <pre> graph LR Admin[Administrator] --- SMA((Social Media Analysis)) SMA -.-> <<Include>> UI((User identification)) SMA -.-> <<Include>> AA((Activity assessment)) SMA -.-> <<Include>> HD((Hashtag detection)) </pre>

TABLE 3 - SCENARIO SCE-PPR-02

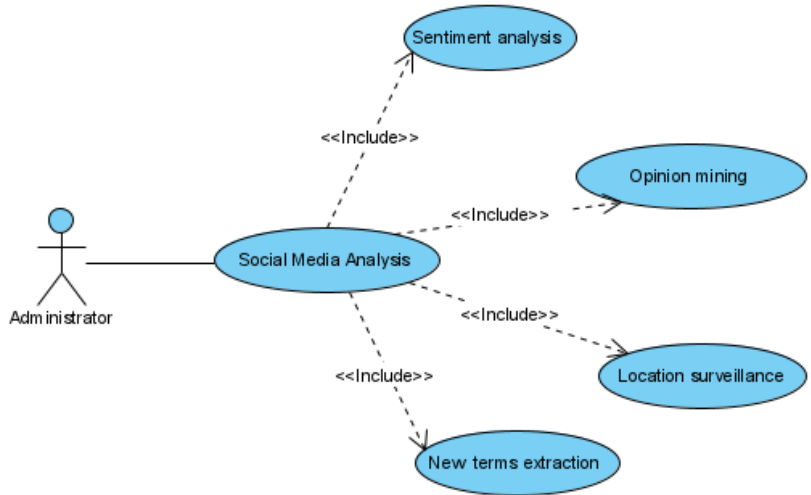
Section	Description
ID	SCE-PPR-03
Title	Assessment of social media activities against perceived radicalisation efforts
Description	Linking of data about terroristic groups and attacks with radicalization efforts
Actors	Sector Analyst, Policy makers
Objectives	<ul style="list-style-type: none"> Observe social networks' posts Identify comments on posts Observe social interactions
Pre-Conditions	<ul style="list-style-type: none"> Identify activities on social media networks Identify events and attitude about radicalization efforts
Process Description	Data and Semantics linking
Variations	N/A
Post-Conditions	N/A
UML User Case Diagram	 <pre> graph LR Admin[Administrator] --- SMA((Social Media Analysis)) SMA -.-> <<Include>> SA((Sentiment analysis)) SMA -.-> <<Include>> OM((Opinion mining)) SMA -.-> <<Include>> LS((Location surveillance)) SMA -.-> <<Include>> NTE((New terms extraction)) </pre> <p>The diagram shows an actor labeled 'Administrator' connected to a central use case 'Social Media Analysis'. This central use case has four dashed arrows pointing to other use cases, each labeled with the stereotype '<<Include>>'. The included use cases are 'Sentiment analysis', 'Opinion mining', 'Location surveillance', and 'New terms extraction'.</p>

TABLE 4 - SCENARIO SCE-PPR-03

Section	Description
ID	SCE-PPR-04
Title	Open Datasets Configuration
Description	Insert new keywords to search
Actors	Administrator, Policy makers
Objectives	Describing elements from open datasets where information has to be analysed
Pre-Conditions	<ul style="list-style-type: none"> The Administrator should have been identified on the platform

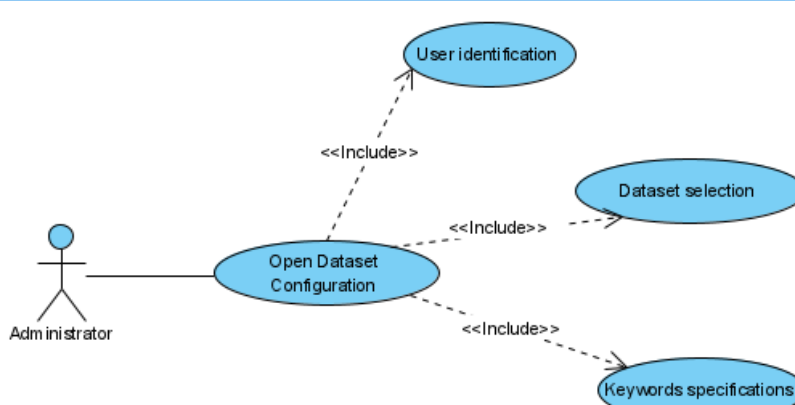
Section	Description
	<ul style="list-style-type: none"> Datasets used - GTD² or RDWTI³
Process Description	<ul style="list-style-type: none"> Administrator selects “My research interests” Systems shows a wizard where the user enters configuration parameters - topics to inspect Administrator inserts, modifies or deletes keywords System saves configuration and adds new parameters to be used by the probe
Variations	N/A
Post-Conditions	<ul style="list-style-type: none"> Text is identified and analysed
UML User Case Diagram	 <pre> graph LR Admin[Administrator] --- ODC([Open Dataset Configuration]) ODC -.-> <<Include>> UI([User identification]) ODC -.-> <<Include>> DS([Dataset selection]) ODC -.-> <<Include>> KS([Keywords specifications]) </pre>

TABLE 5 - SCENARIO SCE-PPR-04

Section	Description
ID	SCE-PPR-06
Title	Open Datasets Analysis
Description	Identify radicalization efforts and the actors (individual, groups) involved
Actors	Data Collector, Analyst, Policy makers
Objectives	<ul style="list-style-type: none"> Identify individuals and organised groups based on several observations Identify new keywords
Pre-Conditions	Verify activity on the open datasets
Process Description	<ul style="list-style-type: none"> System read the text to process System analysed and categorized the text
Variations	N/A

² <https://www.start.umd.edu/gtd>

³ <https://www.rand.org/nsrd/projects/terrorism-incidents.html>

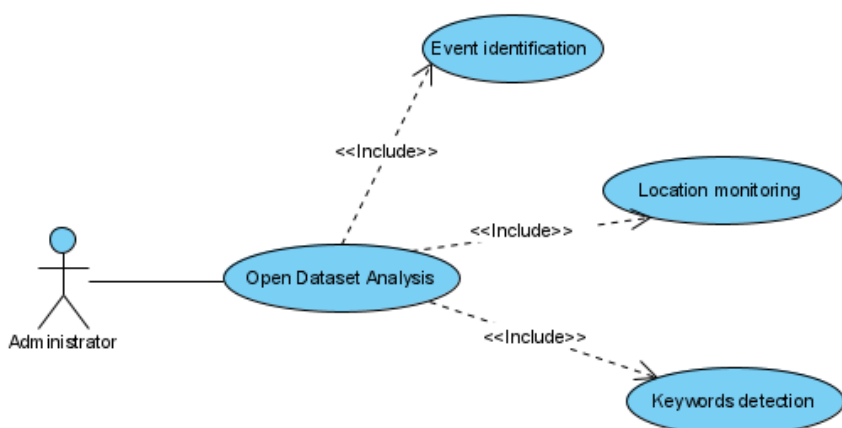
Post-Conditions	<ul style="list-style-type: none"> Text is identified and analysed Text categorization defined (taxonomy or ontology)
UML User Case Diagram	 <pre> graph LR Admin[Administrator] --- ODA([Open Dataset Analysis]) ODA -.-> <<Include>> EI([Event identification]) ODA -.-> <<Include>> LM([Location monitoring]) ODA -.-> <<Include>> KD([Keywords detection]) </pre>

TABLE 6 - SCENARIO SCE-PPR-06

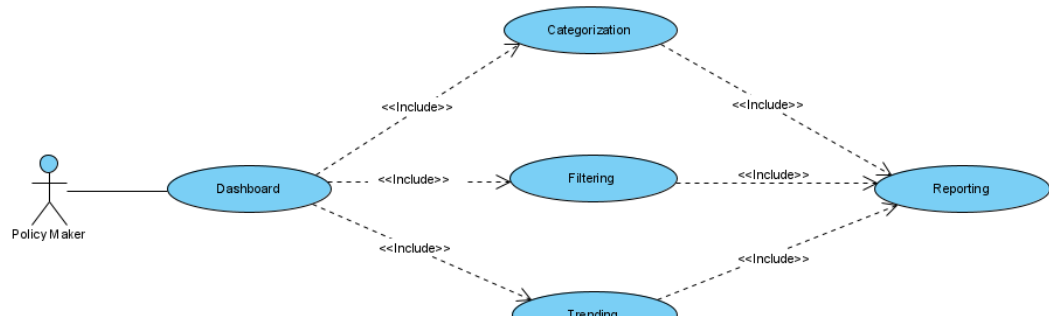
Section	Description
ID	SCE-PPR-08
Title	Visual Representation of Radicalization Trends and Poles
Description	Make policies against radicalization and violent extremism more transparent and open to public scrutiny through the use of the proposed technologies that create and structure open data datasets with statistics about radicalization efforts in social media and provide APIs and visualization tools for accessing them (i.e., making them easily accessible and reusable by third parties)
Actors	Sector Analyst, Policy makers
Objectives	N/A
Pre-Conditions	Ensure continuous interactions and collaboration with relevant stakeholders at any level (local, regional, national and EU)
Process Description	N/A
Variations	N/A
Post-Conditions	N/A
UML User Case Diagram	 <pre> graph LR PM[Policy Maker] --- D([Dashboard]) D -.-> <<Include>> C([Categorization]) D -.-> <<Include>> F([Filtering]) D -.-> <<Include>> T([Trending]) C -.-> <<Include>> R([Reporting]) F -.-> <<Include>> R T -.-> <<Include>> R </pre>

TABLE 7 - SCENARIO SCE-PPR-08

2.1.3 Stakeholder Requirements

The following tables contain the initial list of the stakeholder requirements for the scenarios of this use case that were described in the previous subsection.

Section	Description
ID	REQ- PPR-01
Title	Restricted access
Level of detail	User
Type	FUNC
Description	A username and password are required to configure which information have to be gathered
Additional Information	N/A
Actor	Administrators, End Users
Priority	MAN
Reference Use Case	SCE-PPR-01, SCE-PPR-04
Success Criteria	Nobody without login/password can access to the application
Expected delivery date	M12

TABLE 8 - STAKEHOLDER REQUIREMENT REQ-PPR-01

Section	Description
ID	REQ-PPR-02
Title	Opinion Mining
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks should be analysed in order to better understand what individuals or groups are saying about a specific discussion topic
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-02, SCE-PPR-03, SCE-PPR-05
Success Criteria	Analysis of the opinions using data from social media channels
Expected delivery date	M34

TABLE 9 - STAKEHOLDER REQUIREMENT REQ-PPR-02

Section	Description
ID	REQ-PPR-03

Section	Description
Title	Sentiment Analysis
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks and open datasets should be analysed in order to know how individuals or groups feel about a specific discussion topic and capture their feelings
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-02, SCE-PPR-03, SCE-PPR-05
Success Criteria	Understanding the sentiment using data from social media channels
Expected delivery date	M34

TABLE 10 - STAKEHOLDER REQUIREMENT REQ-PPR-03

Section	Description
ID	REQ-PPR-04
Title	Text classification
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks and open datasets should be grouped in predefined “clusters”
Additional Information	Classification must follow the parameters specified by policy regulators
Actor	Administrator, Policy Makers
Priority	MAN
Reference Use Case	SCE-PPR-02, SCE-PPR-03, SCE-PPR-05
Success Criteria	All texts have to be tagged into one or more clusters
Expected delivery date	M24

TABLE 11 - STAKEHOLDER REQUIREMENT REQ-PPR-04

Section	Description
ID	REQ-PPR-05
Title	Extraction of entities
Level of detail	Stakeholder
Type	FUNC

Description	More relevant entries (age range, location (city area), attack types, etc.) should be extracted from gathered data
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-03, SCE-PPR-05, SCE-PPR-07
Success Criteria	Entities extracted from texts
Expected delivery date	M24

TABLE 12 - STAKEHOLDER REQUIREMENT REQ- PPR-05

Section	Description
ID	REQ-PPR-06
Title	Personal Data
Level of detail	Stakeholder
Type	DATA
Description	Personal data like name, address, will not be stored or, at list, they will be stored after pseudonymization and/or aggregation/anonymization
Additional Information	Full compliance with GPDR and national laws
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-03
Success Criteria	No personal data is linked to the information extracted and stored in the system
Expected delivery date	M12

TABLE 13 - STAKEHOLDER REQUIREMENT REQ- PPR-06

Section	Description
ID	REQ-PPR-07
Title	Twitter information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from tweets - text, user (alias), channel, location, date, origin source
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	MAN

Reference Use Case	SCE-PPR-03
Success Criteria	Data from relevant tweets have been extracted
Expected delivery date	M24

TABLE 14 - STAKEHOLDER REQUIREMENT REQ-PPR-07

Section	Description
ID	REQ-PPR-08
Title	Facebook information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from posts and comments in group discussions - text, user (alias), location, attack type, etc.
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-03
Success Criteria	Data from relevant posts have been extracted
Expected delivery date	M24

TABLE 15 - STAKEHOLDER REQUIREMENT REQ-PPR-08

Section	Description
ID	REQ-PPR-09
Title	Reddit Information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from Reddit posts - text, user (alias), location, date, origin source
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	DES
Reference Use Case	SCE-PPR-03
Success Criteria	Data from relevant blog posts have been extracted
Expected delivery date	M24

TABLE 16 - STAKEHOLDER REQUIREMENT REQ-PPR-09

Section	Description
ID	REQ-PPR-10
Title	RSS & Web Pages
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from web pages - text, source, date, title
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-03
Success Criteria	Data from relevant pages have been extracted
Expected delivery date	M24

TABLE 17: STAKEHOLDER REQUIREMENT REQ-PPR-10

Section	Description
ID	REQ-PPR-11
Title	Information gathered from open datasets
Level of detail	Stakeholder
Type	DATA
Description	Data from different open sources will be gathered
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-05
Success Criteria	Data from relevant open datasets have been extracted
Expected delivery date	M12

TABLE 18 - STAKEHOLDER REQUIREMENT REQ-PPR-11

Section	Description
ID	REQ-PPR-12
Title	Data Analysis in near real-time
Level of detail	Stakeholder
Type	PERF

Description	Information collected from different dataset should be analysed every predefined time
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-PPR-03, SCE-PPR-05
Success Criteria	Analysis of data is available
Expected delivery date	M34

TABLE 19 - STAKEHOLDER REQUIREMENT REQ-PPR-12

Section	Description
ID	REQ-PPR-13
Title	Visualization
Level of detail	Stakeholder
Type	L&F
Description	Dashboard should be able to visualise relevant information from different data sources at a glance
Additional Information	N/A
Actor	End User, Policy Maker
Priority	DES
Reference Use Case	SCE-PPR-06, SCE-PPR-07
Success Criteria	The results are visualised and accessible
Expected delivery date	M34

TABLE 20 - STAKEHOLDER REQUIREMENT REQ-PPR-13

Section	Description
ID	REQ-PPR-14
Title	Risk Prediction
Level of detail	Stakeholder
Type	L&F
Description	Dashboard should show predictions on potential risks/threats and their location
Additional Information	N/A
Actor	End User, Policy Maker
Priority	DES

Reference Use Case	SCE-PPR-06, SCE-PPR-07
Success Criteria	Successful prediction of at least one risk
Expected delivery date	M34

TABLE 21 - STAKEHOLDER REQUIREMENT REQ-PPR-14

Section	Description
ID	REQ-PPR-15
Title	Working Hours
Level of detail	Stakeholder
Type	SUP
Description	<ul style="list-style-type: none"> Gathering and analysis information should be working 24x7 Web page should be available 24x7
Additional Information	N/A
Actor	N/A
Priority	DES
Reference Use Case	SCE-PPR-01
Success Criteria	The system is always accessible
Expected delivery date	M34

TABLE 22 - STAKEHOLDER REQUIREMENT REQ-PPR-15

2.2 UC#2: Intelligent policies for the development of agrifood industry

2.2.1 Goals and Objectives

The main objective of this use case is to foster the use of analytics technologies that will facilitate the implementation of new investment methodologies in the Aragon region with a focus on products with a Protected Designation of Origin (also called in PolicyCLOUD Denomination of Origin).

The most important ideas of the use case are:

- Identification of heterogeneous data sources like social networks (Twitter, Facebook, Instagram), open data sources, and any other documents provided by the government, industries, etc., in order to extract hidden patterns and information.
- Extracting, analysing and classifying information based on defined ontologies to generate reports about the state of the art, and news about the different products covered by the denomination of origins in Aragón.
- Knowing deeply which are the new trends on international markets, the emerging issues about the product covered, as well as the policy recommendation.
- Brand analysis of Aragón's DO and competence, and recommendations to implement policies.
- Creating KPIs to analyse the status of denomination of Origin.
- Creating and implementing new policies to help to create proposals of differential value on the agri-food sector in Aragon which helps to the specialization and development of Denomination of Origin.
- Evaluating the impact of implemented policies and comparison with the older ones.

Taking into account these objectives, the purpose of the use case is to create a tool that allows governments to analyse what is happening and design and improve policies around the Denomination of Origin in Aragon.

2.2.2 Description of Scenarios

The following tables define the various scenarios of this use case.

Section	Description
ID	SCE-IIPDD-01
Title	Configuration Panel
Description	Identifying and configuring data sources to analyse: <ul style="list-style-type: none"> • Social network's users and communities • Keywords to search on social networks • Information provided from news webs and blogs • Configuration of categories to classify information provided from different channels

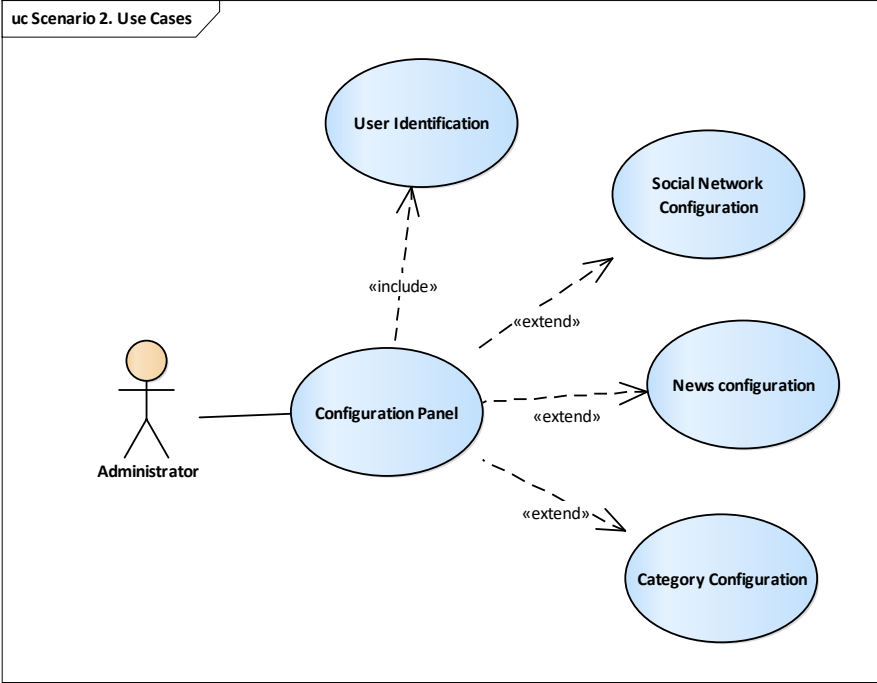
Section	Description
Actors	Administrator, Policy makers
Objectives	Describing elements in order to identify relevant information
Pre-Conditions	<ul style="list-style-type: none"> The Administrator should be identified on the platform Social networks used - Twitter, Facebook, LinkedIn and Instagram Other channels - news pages, Blogs
Process Description	<ul style="list-style-type: none"> <<Include User Identification>> Administrator enters login and password System presents different options to select under the control panel Administrator selects the option to configure System shows configuration options Administrator configure data sources System saves configuration
Variations	<p>1a. If user's login or password is wrong, the user will not be able to access the control panel</p> <p>2a. <<Extends Social Network configuration>></p> <p>2b. <<Extends News Configuration>></p> <p>2c. <<Extends Category Configuration>></p>
Post-Conditions	<ul style="list-style-type: none"> New sources are identified and analysed Hierarchical text categorization defined (taxonomy or ontology)
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR Admin[Administrator] --- CP((Configuration Panel)) CP -.-> «include» UI((User Identification)) CP -.-> «extend» SNC((Social Network Configuration)) CP -.-> «extend» NC((News configuration)) CP -.-> «extend» CC((Category Configuration)) </pre>

TABLE 23 - SCENARIO SCE-IIPDD-01

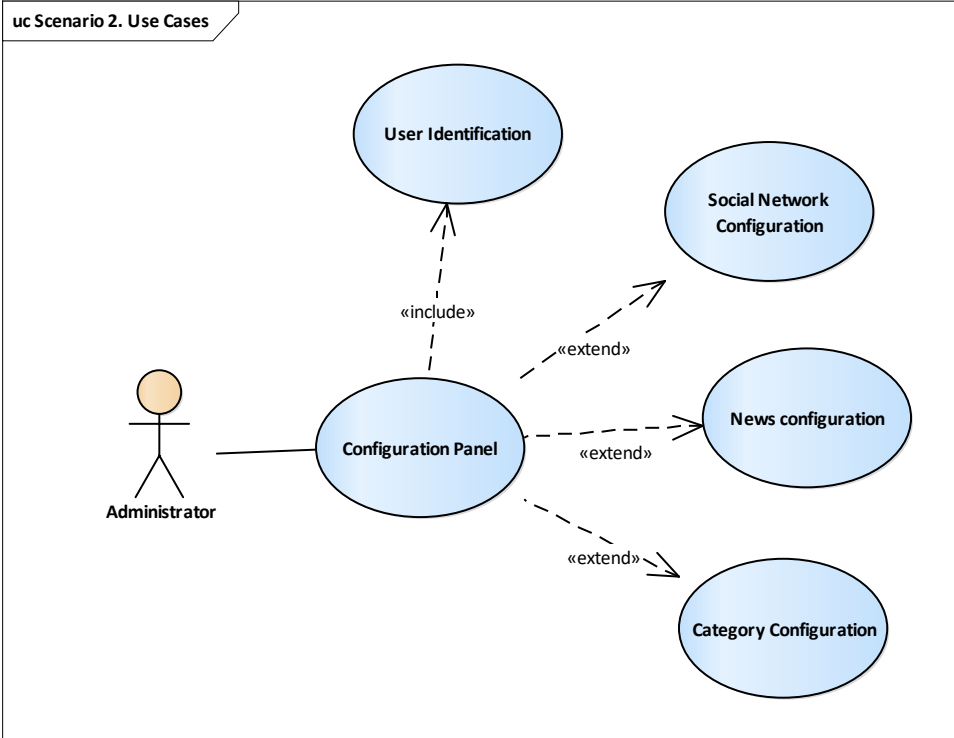
Section	Description
ID	SCE-IIPDD-02
Title	Social Network Configuration
Description	Inserting new users, communities and words to search
Actors	Administrator, Policy makers
Objectives	Describing elements from social networks where information has to be analysed
Pre-Conditions	<ul style="list-style-type: none"> The Administrator should be identified on the platform Social networks used - Twitter, Facebook, LinkedIn and Instagram
Process Description	<ul style="list-style-type: none"> Administrator selects "Social Networks Configuration" System shows a wizard where the administrator enters configuration parameters - social networks, user, or community to follow, terms of ontology to select the topics to inspect from Aragon's Denomination of Origin Administrator inserts, modifies or deletes social network configuration System saves configuration and adds new parameters to be used by the probe
Variations	N/A
Post-Conditions	N/A
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR Admin[Administrator] --- CP((Configuration Panel)) CP -.-> «include» UI((User Identification)) CP -.-> «include» SNC((Social Network Configuration)) CP -.-> «include» NC((News configuration)) CP -.-> «extend» CC((Category Configuration)) CP -.-> «extend» NC </pre>

TABLE 24 - SCENARIO SCE-IIPDD-02

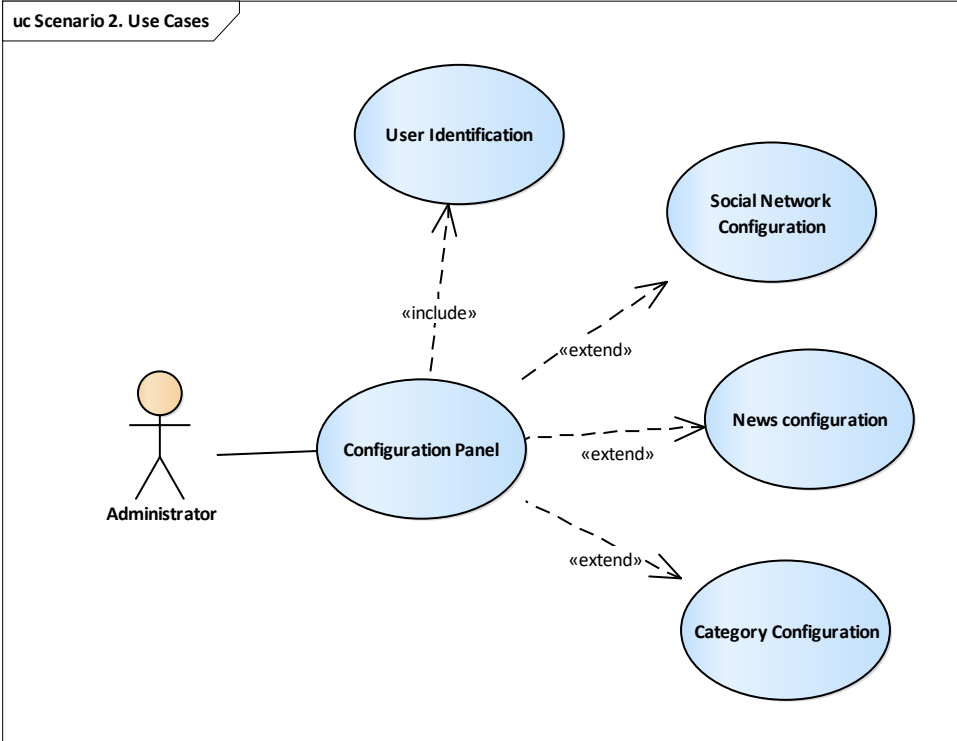
Section	Description
ID	SCE-IIPDD-03
Title	News configuration
Description	Adding new news pages, channels or blogs to extract relevant information
Actors	Administrator, Policy makers
Objectives	Configuration of the system in order to extract relevant information about denominations of origin and related subjects to create and improve policies
Pre-Conditions	<ul style="list-style-type: none"> The Administrator should be identified on the platform Text channels - news pages, Blogs
Process Description	<ul style="list-style-type: none"> Administrator selects “News configuration” System shows a configuration panel where the user defines the source of the document and the type System saves configuration and adds new parameters to be used by the probe
Variations	N/A
Post-Conditions	N/A
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR Admin[Administrator] --- CP([Configuration Panel]) CP -.-> «include» UI([User Identification]) CP -.-> «extend» SNC([Social Network Configuration]) CP -.-> «extend» NC([News configuration]) CP -.-> «extend» CC([Category Configuration]) NC -.-> SNC </pre>

TABLE 25 - SCENARIO SCE-IIPDD-03

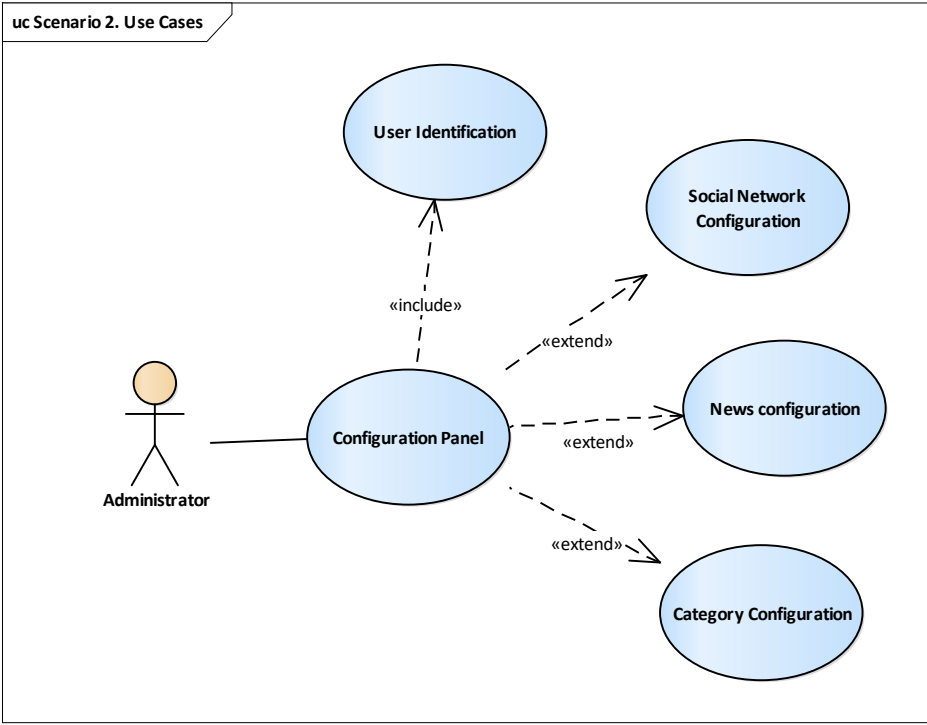
Section	Description
ID	SCE-IIPDD-04
Title	Category Configuration
Description	Definition of a set of hierarchical categories that allow the system to classify information taken from different sources
Actors	Administrator, Policy makers
Objectives	Defining a category for the use case
Pre-Conditions	<ul style="list-style-type: none"> The Administrator should be identified on the platform
Process Description	<ul style="list-style-type: none"> Administrator selects “Category Configuration” Systems shows a wizard where the administrator defines a category and the elements around it Administrator inserts, modifies or deletes information related to categories System saves configuration and adds new parameters to be used by analysers
Variations	N/A
Post-Conditions	<ul style="list-style-type: none"> New categories and words are included in the system
UML User Case Diagram	 <pre> graph LR Admin[Administrator] --- CP((Configuration Panel)) CP -.-> «include» UI((User Identification)) CP -.-> «extend» SNC((Social Network Configuration)) CP -.-> «extend» NC((News configuration)) CP -.-> «extend» CC((Category Configuration)) </pre>

TABLE 26 - SCENARIO SCE-IIPDD-04

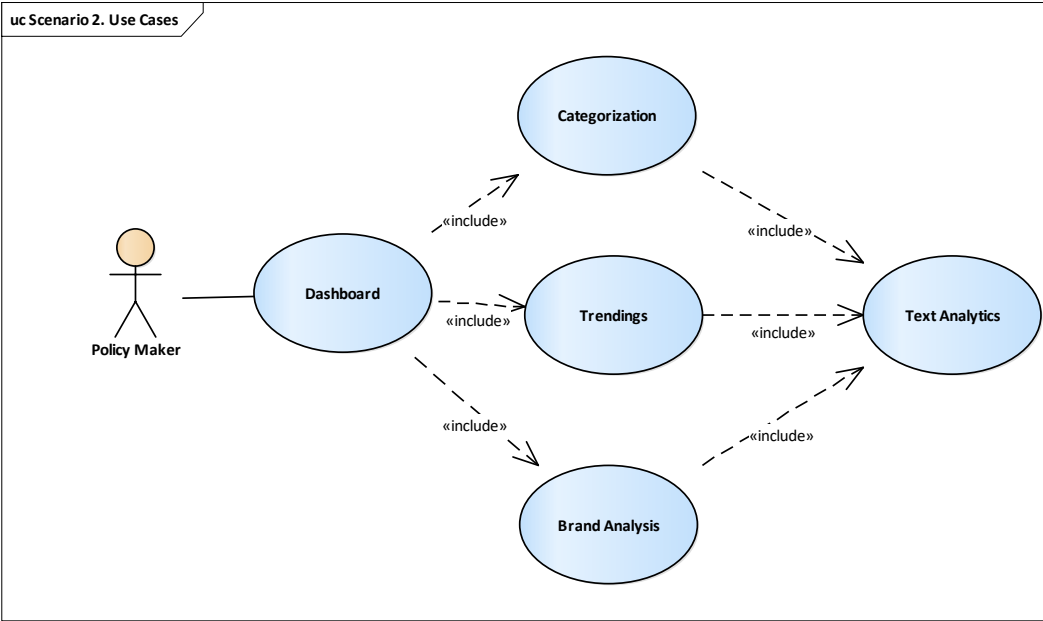
Section	Description
ID	SCE-IIPDD-05
Title	Dashboard
Description	Displaying information which helps policy makers to take the right decision that is related to Denomination of Origin subjects
Actors	Policy Maker, End user
Objectives	Providing support to policy makers
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> System presents different options to select under the dashboard <ul style="list-style-type: none"> Period of time to evaluate Designation of Origin Producer or winery Brand or wine selection Properties of the product (Wine's color) User selects the option to configure System shows a panel with obtained results and recommendations to the user <ul style="list-style-type: none"> Sentimental analysis of the examined sources Geographical map to show where opinions come from Sentimental analysis along the time
Variations	2a. Categorization 2b. Trending 2c. Brand analysis
Post-Conditions	N/A
UML User Case Diagram	 <pre> graph LR PM[Policy Maker] --- Dashboard Dashboard -.-> «include» Categorization Dashboard -.-> «include» Trendings Dashboard -.-> «include» BrandAnalysis[Brand Analysis] Categorization -.-> «include» TextAnalytics[Text Analytics] Trendings -.-> «include» TextAnalytics BrandAnalysis -.-> «include» TextAnalytics </pre>

TABLE 27 - SCENARIO SCE-IIPDD-05

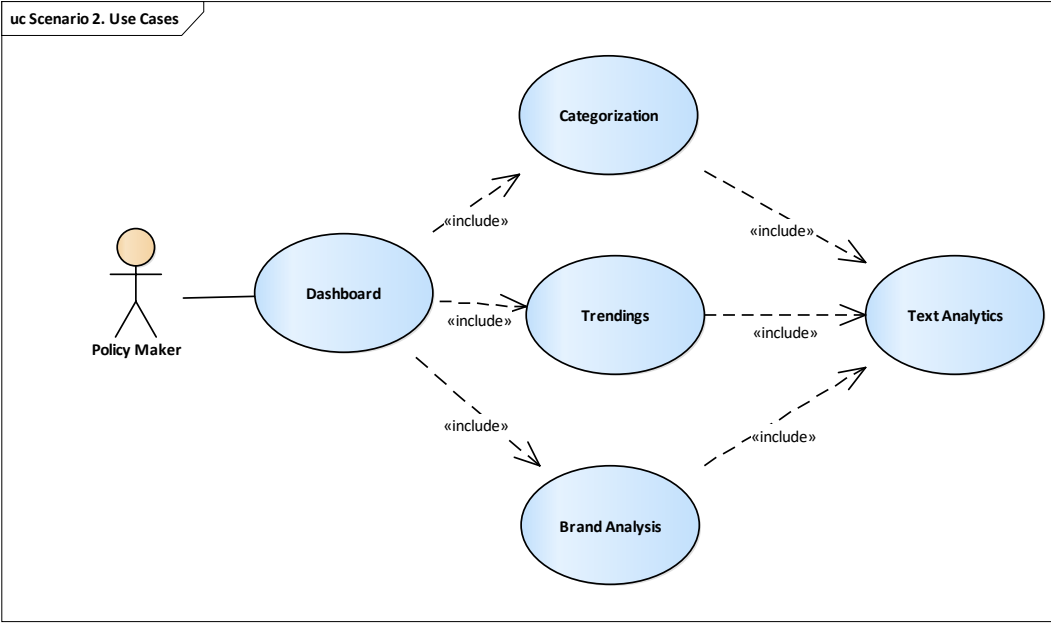
Section	Description
ID	SCE-IIPDD-06
Title	Categorization
Description	Information extracted from social networks, new pages and blogs is classified based on the categories defined by the Administrators
Actors	Policy Makers, End user
Objectives	Classify information into defined categories for report generation
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> User selects "Categorization" System gathers stored information <<Include <i>Text Analytics</i>>> Stored information is analysed and categorized System shows information to the users - Popular categories, categorized texts, etc. User uses filters to select which information to analyse System recalculates which information to show
Variations	N/A
Post-Conditions	Recommendations to the user are shown based on categorization
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR PM[Policy Maker] --- Dashboard Dashboard -.-> «include» Categorization Dashboard -.-> «include» Trendings Dashboard -.-> «include» BrandAnalysis[Brand Analysis] Categorization -.-> «include» TextAnalytics[Text Analytics] Trendings -.-> «include» TextAnalytics BrandAnalysis -.-> «include» TextAnalytics </pre>

TABLE 28 - SCENARIO SCE-IIPDD-06

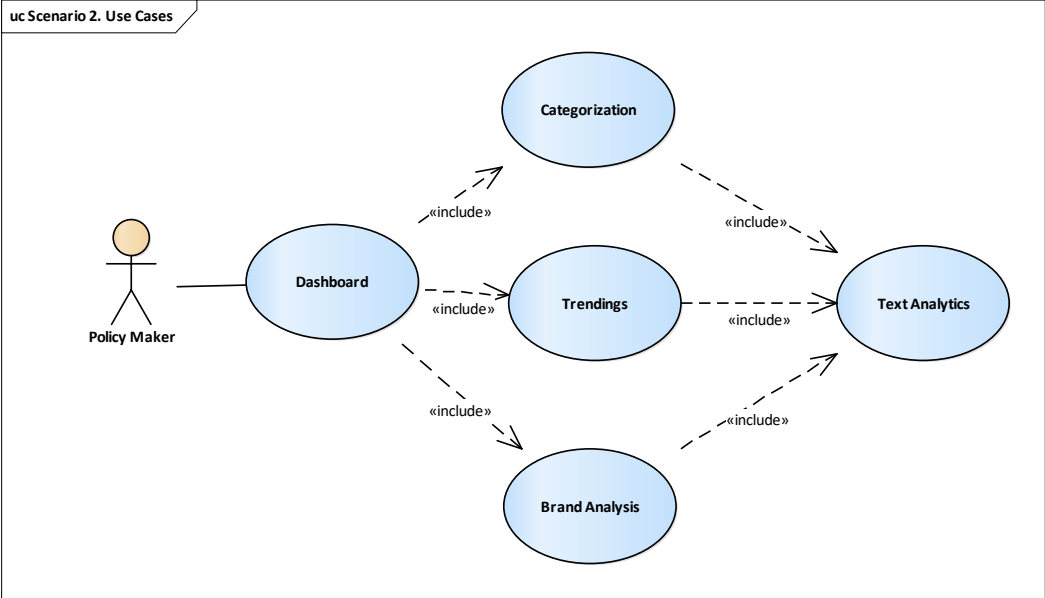
Section	Description
ID	SCE-IIPDD-07
Title	Trendings
Description	Information extracted from social networks, webpages and blogs is analysed in order to discover new trends and generate new recommendations to the user based on opinion analysis
Actors	Policy Makers, End user
Objectives	Discover hidden patterns and new tendencies, make recommendations
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> • User selects "Trending" • System gathers stored information • <<Include <i>Text Analytics</i>>> Stored information is analysed (entities identification, opinion analysis) • System shows information to the users • User uses filters to select which information to analyse • System recalculates which information to show
Variations	N/A
Post-Conditions	Recommendations to the user are shown based on discovered patterns and tendencies
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR PM[Policy Maker] --- Dashboard Dashboard -.-> «include» Categorization Dashboard -.-> «include» Trendings Dashboard -.-> «include» BrandAnalysis[Brand Analysis] Categorization -.-> «include» TextAnalytics[Text Analytics] Trendings -.-> «include» TextAnalytics BrandAnalysis -.-> «include» TextAnalytics </pre>

TABLE 29 - SCENARIO SCE-IIPDD-07

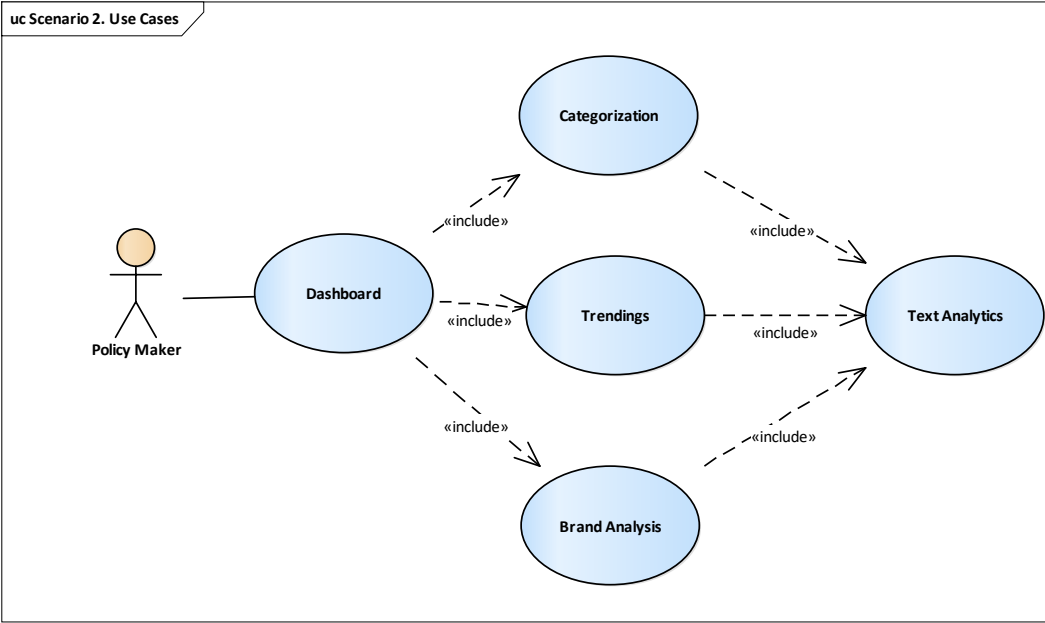
Section	Description
ID	SCE-IIPDD-08
Title	Brand Analysis
Description	Information extracted from social networks, RSS channels and blogs is analysed in order to analyse the wine market
Actors	Policy Makers, End user
Objectives	Study the market in order to create new policies or adapt the existing ones
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> • User selects “Brand analysis” • System gathers stored information • <<Include <i>Text Analytics</i>>> Stored information is analysed • System shows information to the users - Popular brands, new tendencies related to the wine market • User selects use filters to select which information to analyse • System recalculates which information to show
Variations	N/A
Post-Conditions	Recommendations to the user are shown based on brand analysis
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p>  <pre> graph LR PM[Policy Maker] --- Dashboard Dashboard -.-> «include» Categorization Dashboard -.-> «include» Trendings Dashboard -.-> «include» BrandAnalysis[Brand Analysis] Categorization -.-> «include» TextAnalytics[Text Analytics] Trendings -.-> «include» TextAnalytics BrandAnalysis -.-> «include» TextAnalytics </pre>

TABLE 30 - SCENARIO SCE-IIPDD-08

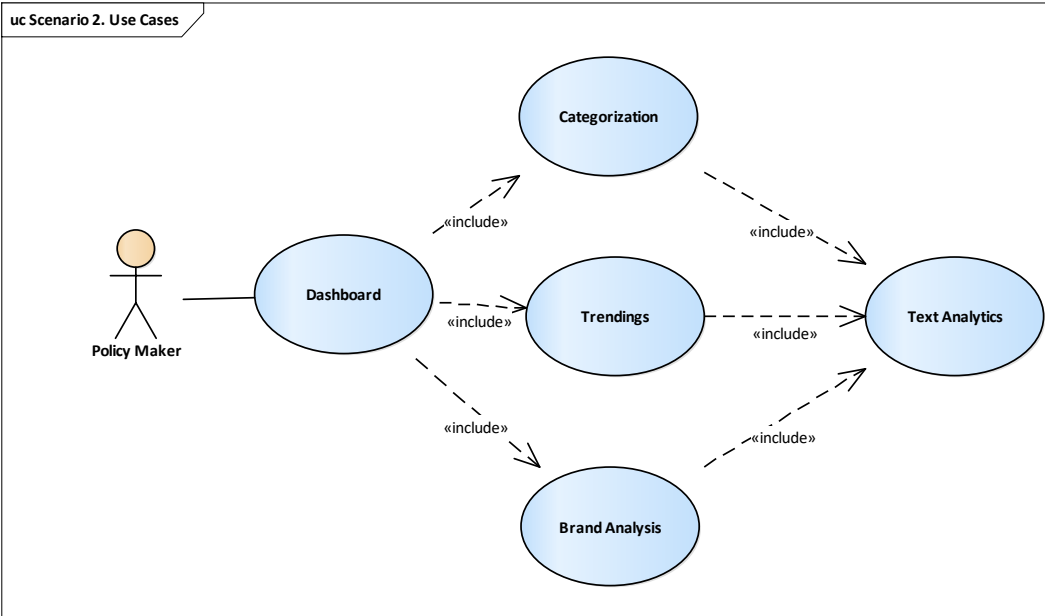
Section	Description
ID	SCE-IIPDD-09
Title	Text Analysis
Description	Information provided by social networks, RSS channels and blogs is analysed and classified
Actors	N/A
Objectives	Opinion Analysis, Sentiment Analysis, extraction of entities
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> System reads the text to process System analyses and categorizes the text
Variations	N/A
Post-Conditions	N/A
UML User Case Diagram	 <pre> graph LR PM[Policy Maker] --- Dashboard Dashboard -.-> «include» Categorization Dashboard -.-> «include» Trendings Dashboard -.-> «include» BrandAnalysis[Brand Analysis] Categorization -.-> «include» TextAnalytics[Text Analytics] Trendings -.-> «include» TextAnalytics BrandAnalysis -.-> «include» TextAnalytics </pre>

TABLE 31 - SCENARIO SCE-IIPDD-09

Section	Description
ID	SCE-IIPDD-10
Title	Price analysis capturing
Description	Information provided by wine Ecommerce is crawled in order to generate a repository of temporal data
Actors	N/A
Objectives	Prices analysis
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> System reads eCommerce website System analyses and categorizes the text and prices
Variations	N/A

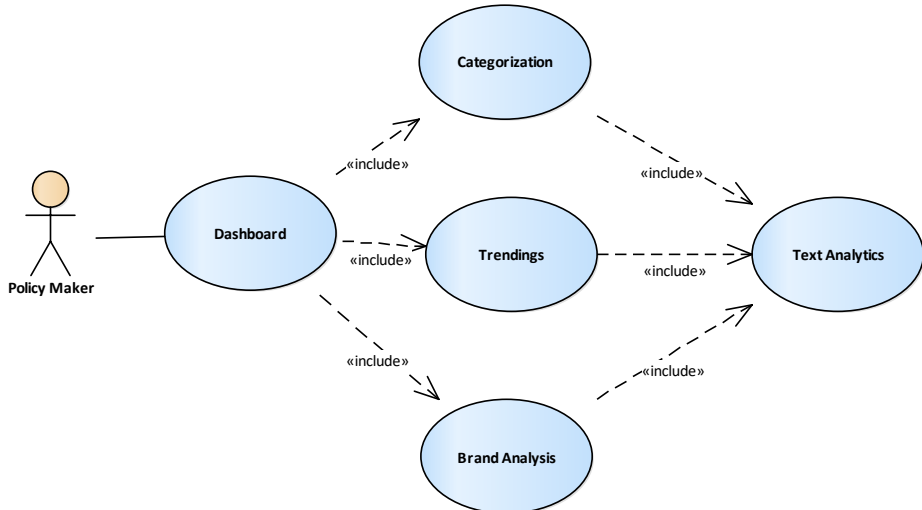
Section	Description
Post-Conditions	N/A
UML User Case Diagram	<p>uc Scenario 2. Use Cases</p> 

TABLE 32 - SCENARIO SCE-IIPDD-10

Section	Description
ID	SCE-IIPDD-11
Title	Price analysis Visualization and alarm generation
Description	Wine prices are analysed and compared, and based on pattern recognition alarms will be generated, whereas the prices will be visualised
Actors	N/A
Objectives	Prices analysis
Pre-Conditions	N/A
Process Description	<ul style="list-style-type: none"> System reads eCommerce data on real tam System makes a visualization of the information
Variations	N/A
Post-Conditions	N/A

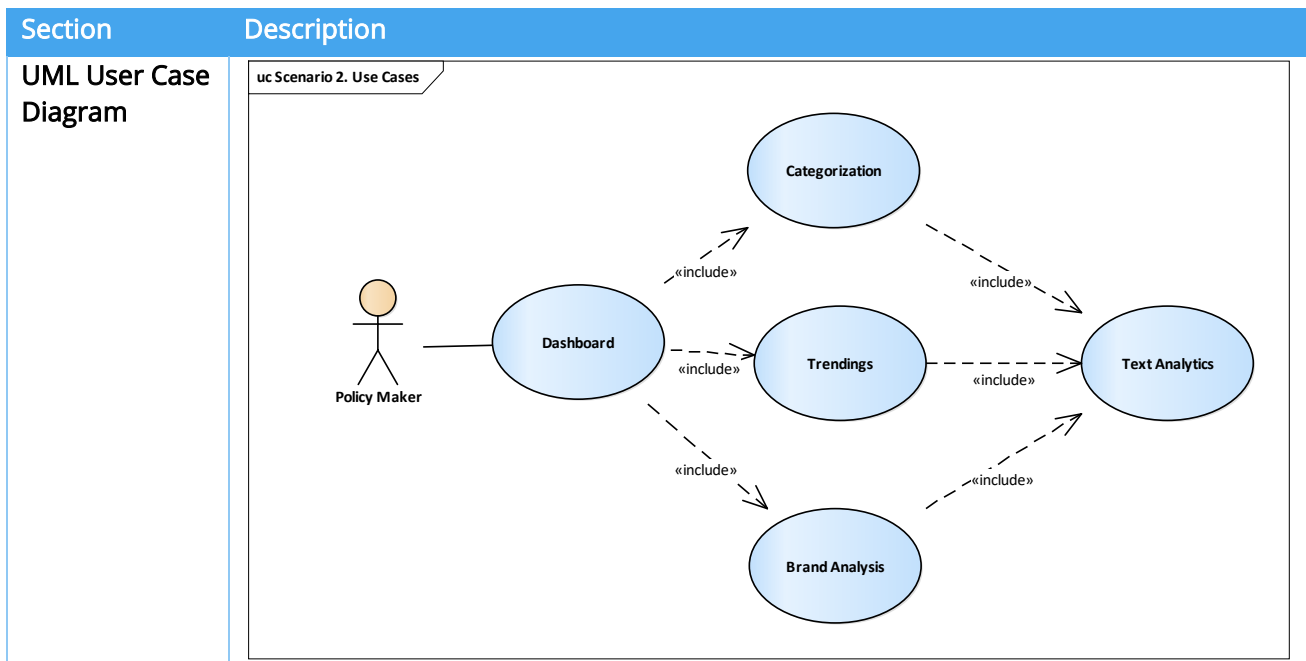


TABLE 33 - SCENARIO SCE-IIPDD-11

2.2.3 Stakeholder Requirements

The following tables contain the initial list of the stakeholder requirements for the scenarios of this use case that were described in the previous subsection.

Section	Description
ID	REQ- IIPDD -01
Title	Restricted access to configuration panel
Level of detail	Stakeholder
Type	FUNC
Description	It will be required a login and a password to configure which information has to be gathered and hierarchical categorization
Additional Information	N/A
Actor	Administrators
Priority	MAN
Reference Use Case	SCE-IIPDD-01
Success Criteria	Nobody without login/password can access the configuration panel
Expected delivery date	M34

TABLE 34 - STAKEHOLDER REQUIREMENT REQ- IIPDD-01

Section	Description
ID	REQ- IIPDD -02
Title	Opinion Analysis
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks and news channels should be analysed in order to know people's opinions about products and brands analysed by Appellation of origin
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Gathered content has assigned a value that defines the level of opinion analysis
Expected delivery date	M34

TABLE 35 - STAKEHOLDER REQUIREMENT REQ- IIPDD-02

Section	Description
ID	REQ- IIPDD -03
Title	Sentiment Analysis
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks and news channels should be analysed in order to know what people feels about products and brands analysed by Appellation of origin
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Gathered content has assigned a value that defines the level of sentimental analysis
Expected delivery date	M34

TABLE 36 - STAKEHOLDER REQUIREMENT REQ- IIPDD-03

Section	Description
ID	REQ- IIPDD -04
Title	Hierarchical text classification
Level of detail	Stakeholder
Type	FUNC
Description	Information gathered from social networks and news channels should be classified regarding to defined hierarchical classification
Additional Information	Classification must be agreed with policy regulators and appellation of origin
Actor	Administrator, Policy Makers
Priority	MAN
Reference Use Case	SCE-IIPDD-04, SCE-IIPDD-06, SCE-IIPDD-09
Success Criteria	All texts have to be classified into one or more categories
Expected delivery date	M34

TABLE 39 - STAKEHOLDER REQUIREMENT REQ- IIPDD-04

Section	Description
ID	REQ- IIPDD -05
Title	Extraction of entities
Level of detail	Stakeholder
Type	FUNC
Description	More relevant entities (e.g., name, location) should be extracted from gathered texts
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-05, SCE-IIPDD-06, SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Entities extracted from texts
Expected delivery date	M34

TABLE 37 - STAKEHOLDER REQUIREMENT REQ- IIPDD-05

Section	Description
ID	REQ- IIPDD -06
Title	Personal Data
Level of detail	Stakeholder
Type	DATA
Description	Personal data like name, address, will not be stored or, at list, they will be stored after pseudonymization and/or aggregation/anonymization

Additional Information	Taking into account GDPR
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-09
Success Criteria	No personal data is displayed on the dashboard
Expected delivery date	M34

TABLE 38 - STAKEHOLDER REQUIREMENT REQ- IIPDD-06

Section	Description
ID	REQ- IIPDD -07
Title	Twitter information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from tweets - text, location, date), origin source
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Tweets that has been selected are then related with the ontology and stored on the system
Expected delivery date	M34

TABLE 39 - STAKEHOLDER REQUIREMENT REQ- IIPDD-07

Section	Description
ID	REQ- IIPDD -08
Title	Facebook Information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from Facebook posts - text, location, date, origin source
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Facebook posts have been stored correctly and related the ontology in a correct way

Expected delivery date	M34
-------------------------------	-----

TABLE 40 - STAKEHOLDER REQUIREMENT REQ- IIPDD-08

Section	Description
ID	REQ- IIPDD -09
Title	Instagram Information
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from Instagram posts - text, location, date, origin source
Additional Information	Only information for public accounts will be collected
Actor	N/A
Priority	DES
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	Instagram posts are stored correctly and classified depending on the ontology
Expected delivery date	M34

TABLE 41 - STAKEHOLDER REQUIREMENT REQ- IIPDD-09

Section	Description
ID	REQ- IIPDD -11
Title	RSS & Blogs web pages
Level of detail	Stakeholder
Type	DATA
Description	Information to gather from news web pages - text, source, date, title, but it has to be refined in updated versions of the document
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-07, SCE-IIPDD-08, SCE-IIPDD-09
Success Criteria	RSS and blog web pages' content should be stored and classified depending on the ontology's criteria
Expected delivery date	M34

TABLE 42 - STAKEHOLDER REQUIREMENT REQ- IIPDD-11

Section	Description
ID	REQ- IIPDD -12
Title	Information gathering in near real-time

Section	Description
Level of detail	Stakeholder
Type	PERF
Description	Data from different sources should be taken in near real-time
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	N/A
Success Criteria	Gathered information should be stored and classified depending on ontology's criteria
Expected delivery date	M34

TABLE 43 - STAKEHOLDER REQUIREMENT REQ- IIPDD-12

Section	Description
ID	REQ- IIPDD -13
Title	Data Analysis in near real-time
Level of detail	Stakeholder
Type	PERF
Description	Information collected from different dataset should be analysed immediately
Additional Information	N/A
Actor	N/A
Priority	MAN
Reference Use Case	SCE-IIPDD-09
Success Criteria	Information can be analysed with very little delay
Expected delivery date	M34

TABLE 44 - STAKEHOLDER REQUIREMENT REQ- IIPDD-13

Section	Description
ID	REQ- IIPDD -14
Title	Dashboard displaying
Level of detail	Stakeholder
Type	L&F
Description	Dashboard should show more relevant information at a glance
Additional Information	N/A
Actor	End User, Policy Maker

Priority	DES
Reference Use Case	SCE-IIPDD-05, SCE-IIPDD-06, SCE-IIPDD-07, SCE-IIPDD-08
Success Criteria	Dashboard shows analysis of different data sources in a graphical way (sentimental analysis, opinion analysis, geolocation of data, trends). A zone to filter information by date, designation of origin, producer, brand must exist.
Expected delivery date	M34

TABLE 45 - STAKEHOLDER REQUIREMENT REQ- IIPDD-14

Section	Description
ID	REQ- IIPDD -15
Title	Web design
Level of detail	Stakeholder
Type	L&F
Description	Information should be displayed on a web page
Additional Information	N/A
Actor	End User, Policy makers
Priority	MAN
Reference Use Case	SCE-IIPDD-01, SCE-IIPDD-05, SCE-IIPDD-06, SCE-IIPDD-07, SCE-IIPDD-08
Success Criteria	Gathered information about the web during co-creation and evaluation meetings is properly displayed
Expected delivery date	M34

TABLE 46 - STAKEHOLDER REQUIREMENT REQ- IIPDD-15

Section	Description
ID	REQ- IIPDD -16
Title	Level AA Standards ⁴
Level of detail	Stakeholder
Type	USE
Description	Web page should accomplish AA standards of accessibility
Additional Information	N/A
Actor	End User, Policy makers

⁴ <https://www.w3.org/WAI/WCAG2AA-Conformance>

Priority	DES
Reference Use Case	SCE-IIPDD-01, SCE-IIPDD-05, SCE-IIPDD-06, SCE-IIPDD-07, SCE-IIPDD-08
Success Criteria	Test about accessibility should be approved
Expected delivery date	M34

TABLE 47 - STAKEHOLDER REQUIREMENT REQ- IIPDD-16

Section	Description
ID	REQ- IIPDD -17
Title	Mobile platforms
Level of detail	Stakeholder
Type	USE
Description	Designed platform should be accessible through mobile devices
Additional Information	N/A
Actor	End User, Policy makers
Priority	DES
Reference Use Case	SCE-IIPDD-01, SCE-IIPDD-05, SCE-IIPDD-06, SCE-IIPDD-07, SCE-IIPDD-08
Success Criteria	Test about accessibility should be approved
Expected delivery date	M34

TABLE 48 - STAKEHOLDER REQUIREMENT REQ- IIPDD-17

2.3 UC#3 - Facilitating urban policy making and monitoring through crowdsourcing data analysis

2.3.1 Goals and Objectives

The aim of this use case is to **support Sofia Municipality's policy making in important areas of citizen's areas of everyday life**. By improving the policy making in these areas, **the overall quality of citizen's life will be improved**, which is the overall goal of this project.

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

The described scenarios include main policy making areas, aimed to improve in the context of PolicyCLOUD, the following:

- road infrastructure
- air quality
- transport
- parking
- waste collection and waste distribution
- violation of public order

2.3.2 Description of Scenarios

Sofia municipality is constantly working to improve the urban environment and meet the challenges that the city is facing. Evidence-based policy making is crucial for addressing urban challenges in a cost-efficient way, however there is yet no established process to incorporate data into policy making. The PolicyCLOUD project will support Sofia municipality to address this challenge by assessing and validating its policies, considering analytics' results that combine information (big data) of sectors, related to a) transport, parking and road infrastructure; b) waste collection and waste disposal; c) cleanliness of public spaces; d) ecology, green systems; e) violation of public order; and other areas that are of importance to citizens. Proposed technologies will offer the advantage to interpret, manage and analyse big amounts of data both from existing data sources (citizens' Contact Centre) and from new data sources (open datasets that will become available).

The use case scenario will address policies and initiatives in the mentioned domains, where PolicyCLOUD big data analytics technologies can help towards the effective and efficient planning and allocation of resources, identifying hazardous factors and risk situations.

The source of the data is the citizens' Contact Centre, which is operational since 2014 and it is a unique point of direct communication with citizens, industries and institutions to report non-urgent alerts on deviations from normal urban environment. Citizens can file signals for waste collection and disposal, road and traffic problems, general public infrastructure, ecology, public spaces (playgrounds, public gardens and parks), etc. In 2020, mobile application for reporting issues and registration of signals was launched and added to the existing channels of communication through phone and web platform.

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

Section	Description
ID	SCE-RLIMP-01
Title	Transport
Description	<p>Transport of Sofia is a very complex network. New business and urban areas are being constantly developed, and adjustments to the current transport schemes are done. There are 200 smart crossroads equipped with cameras and sensors. The city Centre of Traffic Control is monitoring and managing the data to ensure no delays in public transportation. Periodically, to tackle air pollution problem and incentivize citizens to use the public transport, Sofia Municipality is introducing daily "Green Ticket" for a substantially lower daily price of EUR 0.5 for the whole network, other actions include offering free parking slots in the buffer parkings to stimulate the use of the metro as a means of transport. Also, Sofia Municipality is gradually renovating the vehicle fleet.</p> <p>By using signals from Contact Centre (and potentially data from the Centre of Traffic Control and the Urban Mobility Centre) potential structural changes and improvement of Sofia Municipality transport policies can be analysed before implementation as well as current policies can be assessed.</p>
Actors	End-users, Administrators
Objectives	<ul style="list-style-type: none"> • Improve quality of service • Improve transport times and better connections for citizens • Assess multimodal pricing schemes and initiatives such as "Green ticket"
Pre-Conditions	<ul style="list-style-type: none"> • Data is available from contact center
Process Description	<ul style="list-style-type: none"> • Municipality submits collected data for analysis • PolicyCLOUD technological providers analyse the data • Municipality gets results from analyses • Municipality improves policy making
Variations	N/A

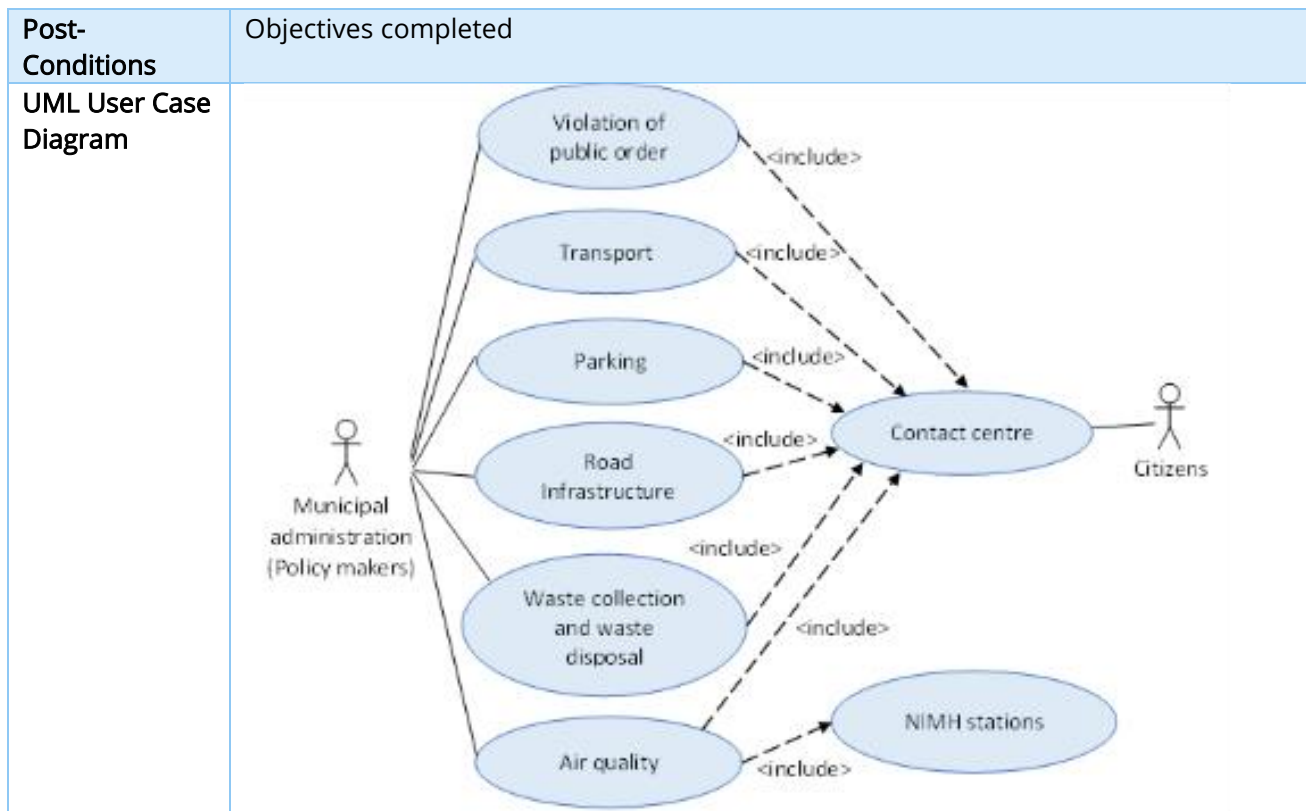


TABLE 49 - SCENARIO SCE-RLIMP-01

Section	Description
ID	SCE-RLIMP-02
Title	Parking
Description	Parking in Sofia is a complex issue with multiple providers. There are a number of municipal parking lots and municipal paid zones within the centre and many private providers. The aim is to optimize provision of parking services for the citizens.
Actors	End-users, Administrator
Objectives	<ul style="list-style-type: none"> Adopt quantity measures for better parking management Improve overall parking capabilities
Pre-Conditions	<ul style="list-style-type: none"> Data is available from a local contact center
Process Description	<ul style="list-style-type: none"> Municipality submits collected data for analysis PolicyCLOUD technological providers analyse the data Municipality gets results from analyses Municipality improves policy making
Variations	N/A
Post-Conditions	Objectives completed

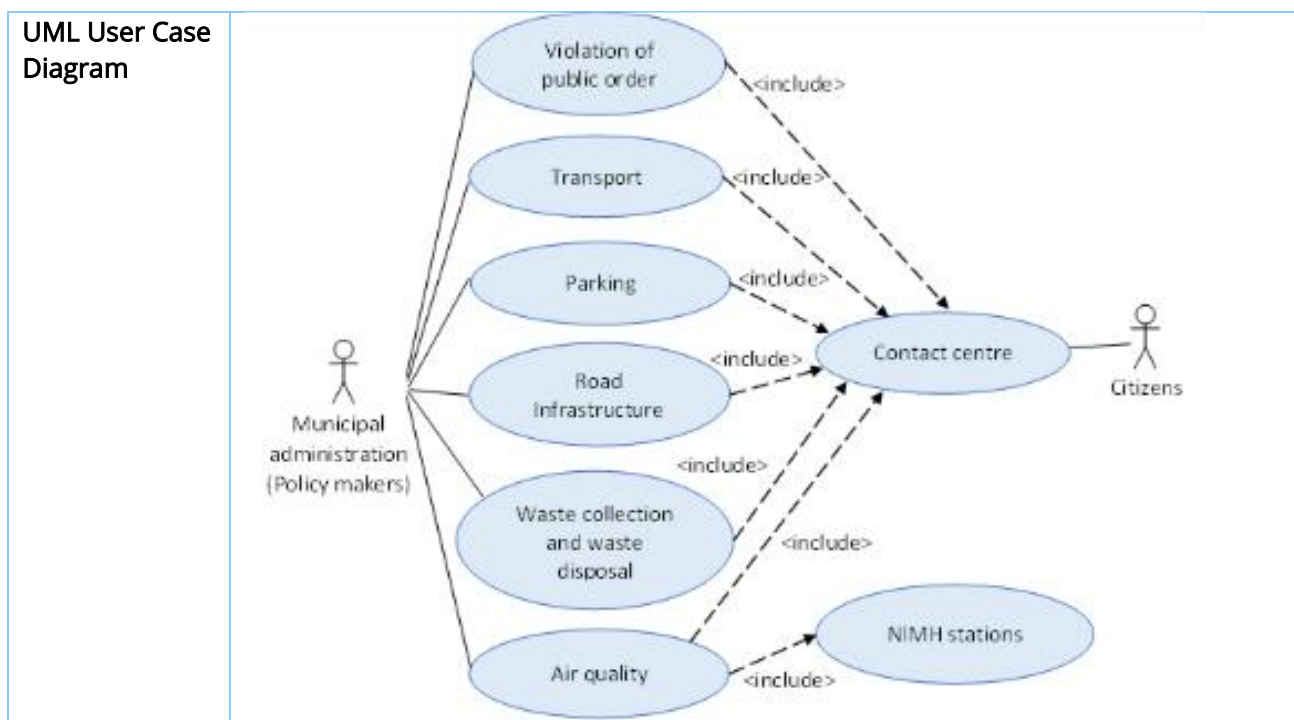


TABLE 50 - SCENARIO SCE-RLIMP-02

Section	Description
ID	SCE-RLIMP-03
Title	Road Infrastructure
Description	Road infrastructure is one of the most important and budget consuming element from the urban environment that impacts citizens' everyday life. Reliable analysis is needed on current situation in all 24 district administrations, in order to foresee and improve long term policy making in the area of read infrastructure.
Actors	End-users, Administrator
Objectives	<ul style="list-style-type: none"> Improving long term policy making in the area of road infrastructure Better envisioning and capacity building of district administrations and municipal administration in solving road infrastructure problems
Pre-Conditions	<ul style="list-style-type: none"> Contact center data
Process Description	<ul style="list-style-type: none"> Municipality submits collected data for analysis PolicyCLOUD technological providers analyse the data Municipality gets results from analyses Municipality improves policy making
Variations	N/A
Post-Conditions	Objectives completed

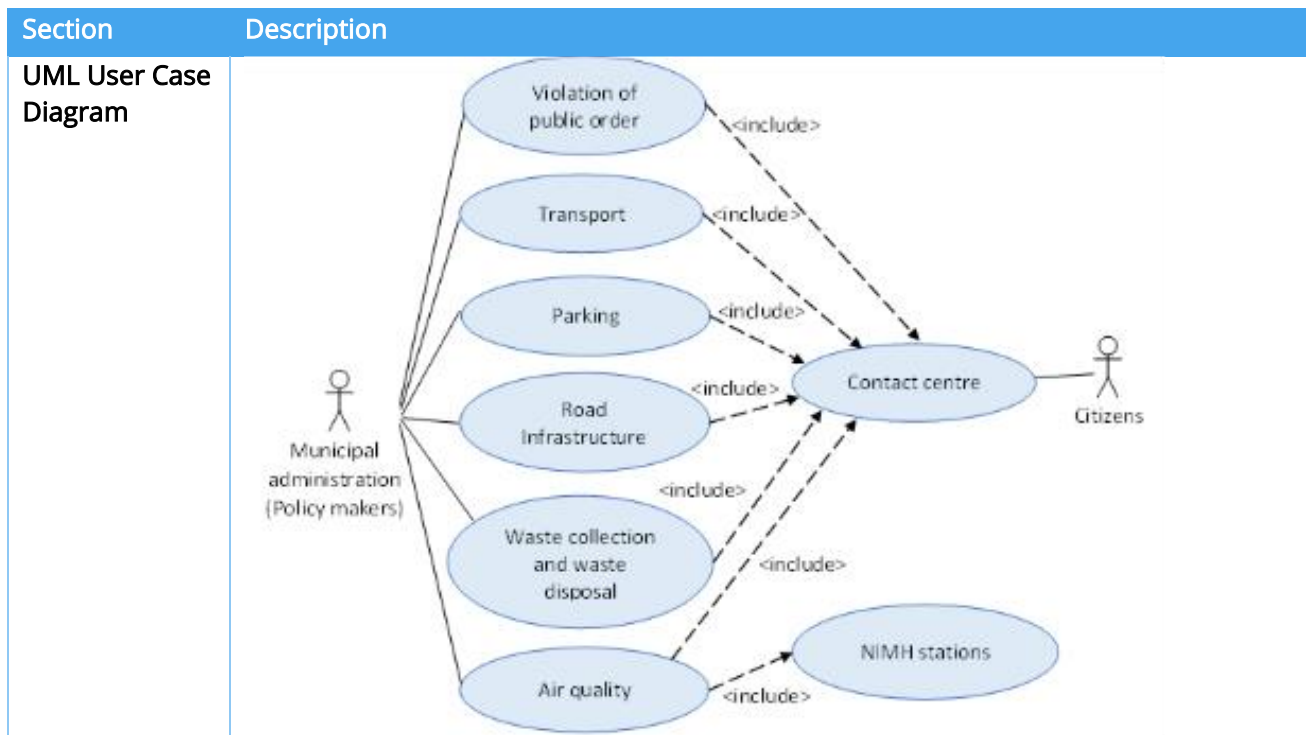


TABLE 51 - SCENARIO SCE-RLIMP-03

Section	Description
ID	SCE-RLIMP-04
Title	Waste collection and waste disposal
Description	Sofia Municipality collects information about waste collection, using smart meters like smart bins, smart garbage trucks, etc. Gathering and analyzing large amounts of data related to waste management, will help Sofia Municipality improve its policy making in the area of urban sustainability and will help the city to become a greener city.
Actors	End-users, Administrator
Objectives	<ul style="list-style-type: none"> • More efficient way of waste collection • Improvement of long term planning and policy making of waste collection and waste disposal using smart meters
Pre-Conditions	<ul style="list-style-type: none"> • Contact center data
Process Description	<ul style="list-style-type: none"> • Municipality submits collected data for analysis • PolicyCLOUD technological providers analyse the data • Municipality gets results from analyses • Municipality improves policy making
Variations	N/A
Post-Conditions	Objectives completed

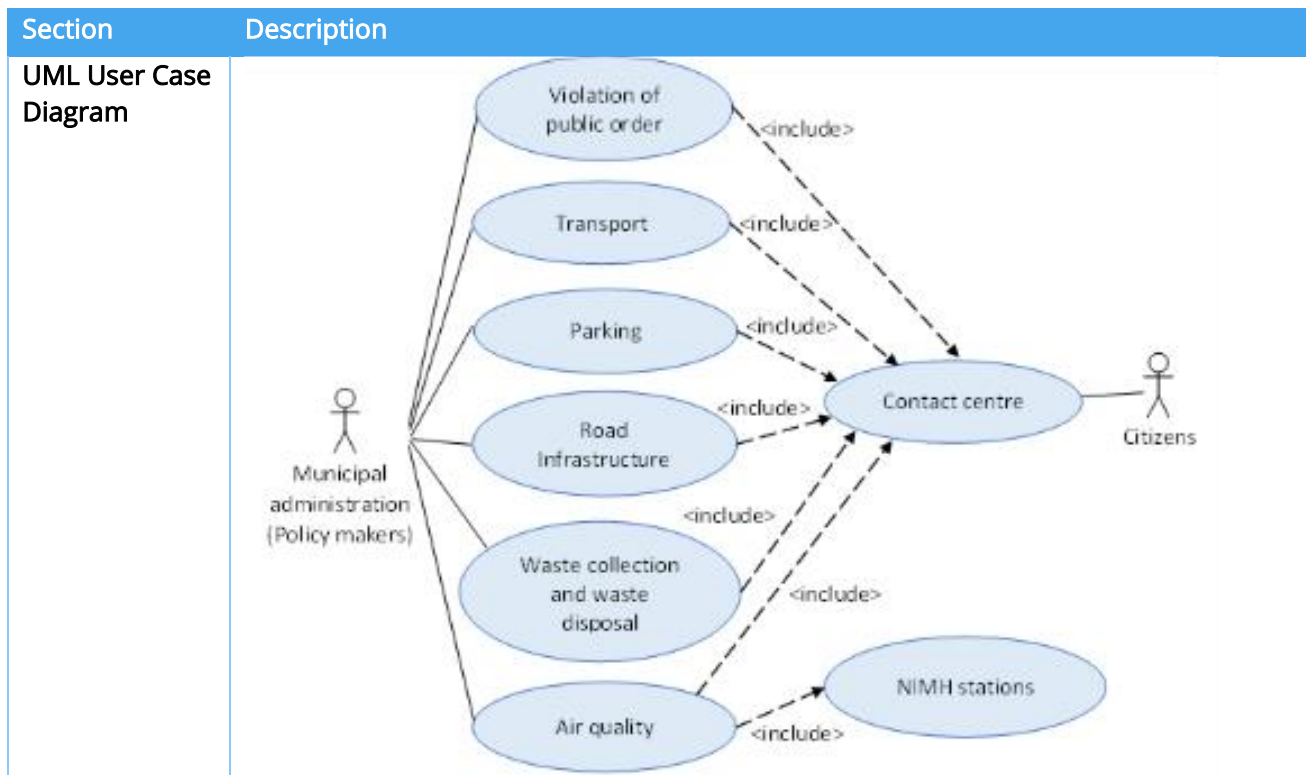


TABLE 52 - SCENARIO SCE-RLIMP-04

Section	Description
ID	SCE-RLIMP-05
Title	Air quality
Description	Under the National Institute for Meteorology and Hydrology NIMH sensors Sofia municipality has built a local network of IoT sensors for monitoring and measurement of the air quality, a smart Internet platform was developed to visualize and store air data. It uses cloud technologies combined with analytical functionality and advanced machine learning opportunities to enable the city administration to take timely action to introduce actions aimed at improving the air quality. The system interface provides the general public with visualization and machine-readable data available through the web. This project provided a system that warns about the danger of high levels of air pollution 48 hours in advance. Executive Environment Agency, which is the owner of the data helps Sofia Municipality, improves its long-term policy making and takes adequate short-term decisions, concerning air quality.
Actors	End-users, Administrator
Objectives	<ul style="list-style-type: none"> Improvement of long-term policy making in the area of air quality
Pre-Conditions	<ul style="list-style-type: none"> Data available from contact center and the NIMH stations
Process Description	<ul style="list-style-type: none"> Municipality submits collected data for analysis PolicyCLOUD technological providers analyse the data Municipality gets results from analyses Municipality improves policy making
Variations	N/A

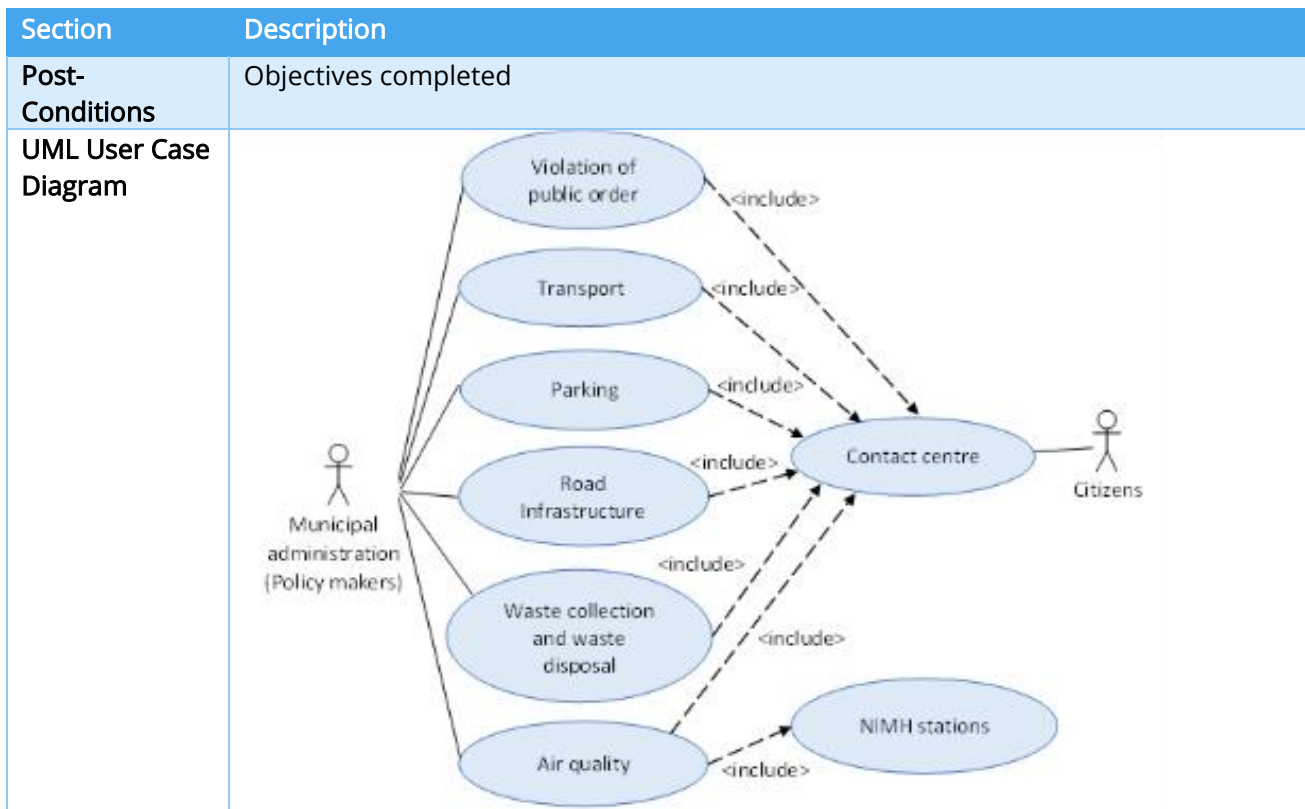


TABLE 53 - SCENARIO SCE-RLIMP-05

Section	Description
ID	SCE-RLIMP-06
Title	Violation of public order
Description	<p>Visualise the signals received via Sofia's Call Centre Call Sofia related to public order violation to provide 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:</p> <ul style="list-style-type: none"> improve long-term policies related to violation prevention and maintaining of public order validate existing policies and investigate if there is a need to update/modify them or create new ones based on the retrieved information
Actors	End-users, Administrator
Objectives	<ul style="list-style-type: none"> Improvement of long-term policy making in the area of public order
Pre-Conditions	<ul style="list-style-type: none"> Data available from contact center
Process Description	<ul style="list-style-type: none"> Municipality submits collected data for analysis PolicyCLOUD technological providers analyse the data Municipality gets results from analyses Municipality improves policy making
Variations	N/A
Post-Conditions	Objectives completed

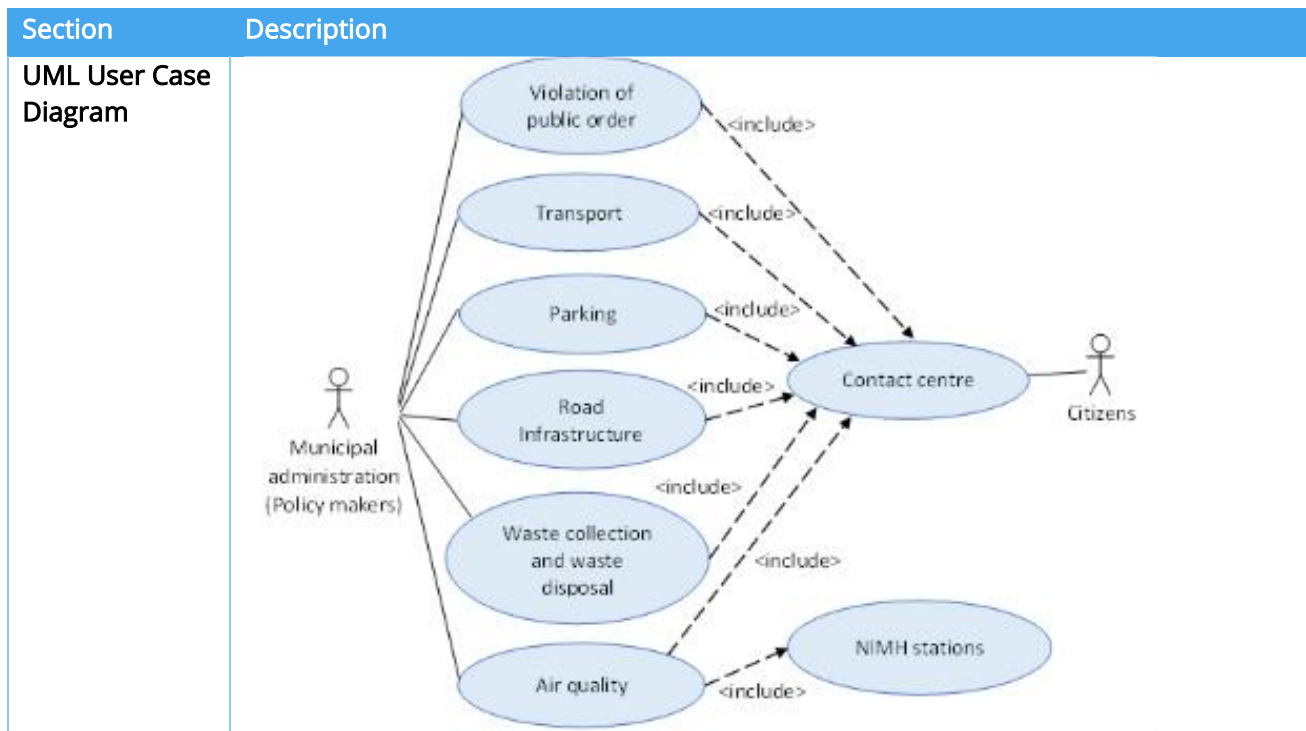


TABLE 54 - SCENARIO SCE-RLIMP-06

2.3.3 Stakeholder Requirements

The following tables contain the initial list of the stakeholder requirements for the scenarios of this use case that were described in the previous subsection.

Section	Description
ID	REQ- RLIMP-01
Title	Provide assessment and visualization
Level of detail	Stakeholder
Type	FUNC
Description	For any policy scenario, the system should be able to predict a set of outcomes
Additional Information	N/A
Actor	End-users, Policy Makers
Priority	MAN for the analysis, DES for visualization
Reference Use Case	SCE-RLIMP-01, SCE-RLIMP-02, SCE-RLIMP-03, SCE-RLIMP-04, SCE-RLIMP-05
Success Criteria	Useful assessment and visualization provided for all six scenarios
Expected delivery date	End of the project M36

TABLE 55 - STAKEHOLDER REQUIREMENT REQ- RLIMP-01

2.4 UC#4 - Predictive analysis towards unemployment risks identification and policy making

2.4.1 Goals and Objectives

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

2.4.2 Description of Scenarios

The aim of this use case is to assist the London borough of Camden's policy creation in helping to tackle social, financial and economic issues that affect citizens. The policy creation process can be sometimes tedious and very complex, and therefore the additional assistance from PolicyCLOUD would be able to save the Policy creator precious time that will boost their work efficiency by allowing them to focus on tackling other pressing issues.

The PolicyCLOUD visualisation technologies will enable policy makers to identify issues, trends, and policy effects and interactions. The PolicyCLOUD analytics technologies will enable to discover insights and find meaningful explanations about the effects of policies. The London use case consist of two scenario, scenario A and B.

Scenario A, is based on Trend Analysis. KPIs and statistics based on the number of citizens claiming government aid analytics that assist with decision making based on unemployment by exploiting trends in the data that help predict future outcomes.

Scenario B will be based on statistical analysis. The PDT should be able to produce statistics on specific time periods. For example, the unemployment rate is expected to go up during the year 2022 due to hardships caused during the recent pandemic. The statistics recorded against the current year can help to identify the possible unemployment rate if there is second wave of infections the following year.

Finally, the Policy Maker should be able to select and view different visualizations in relation to the following categories: gender, age groups and parts of the borough that are most affected by unemployment.

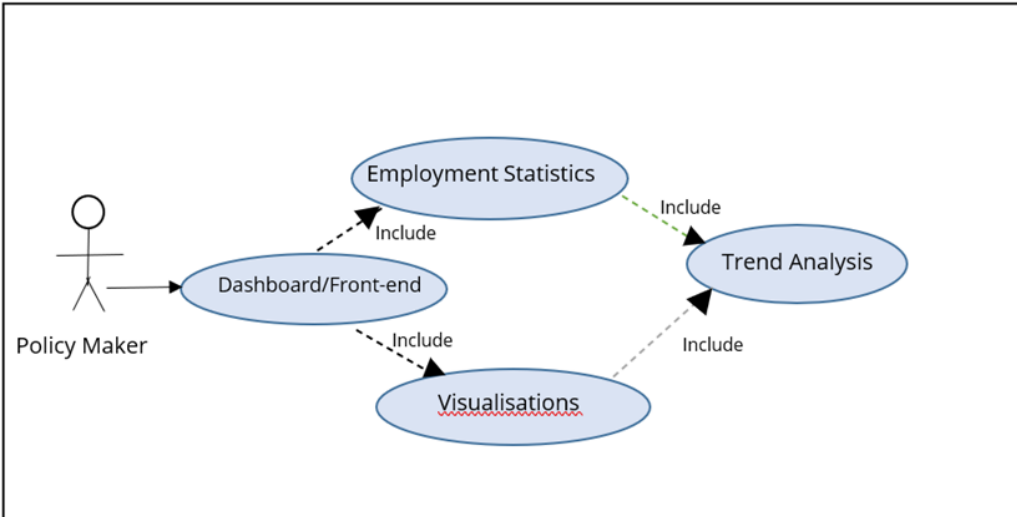
Section	Description
ID	SCE- PAUNRI -01
Title	Analysis capabilities
Description	<p>Linked to scenario A;</p> <ul style="list-style-type: none"> Explore the relation of the data to the underlying population. Create a model to summarize an understanding of how the data relates to the unemployment. Employ predictive analytics to run scenarios that will help guide future actions.
Actors	End User, Policy makers
Objectives	<ul style="list-style-type: none"> Produce clear visualisations that can be interpreted by the policy makers to support the decision-making process Produce line and charts that can demonstrate the statistics for different age groups and male/female citizens that are unemployed Ability to highlight trends via the produced visualisations Support trend analysis
Pre-Conditions	Visualizations related to data available from the front end of the PDT
Process Description	<ul style="list-style-type: none"> London borough of Camden uploads collected data PolicyCLOUD technological providers analyse the data Policy makers get results from scenario analysis London borough benefits from improved policy making process
Variations	N/A
Post-Conditions	Objectives completed
UML User Case Diagram	 <pre> graph LR PM[Policy Maker] --> DF[Dashboard/Front-end] DF -.-> Include ES[Employment Statistics] DF -.-> Include V[Visualisations] ES -.-> Include TA[Trend Analysis] V -.-> Include TA </pre>

TABLE 56 - SCENARIO SCE-PAUNRI-01

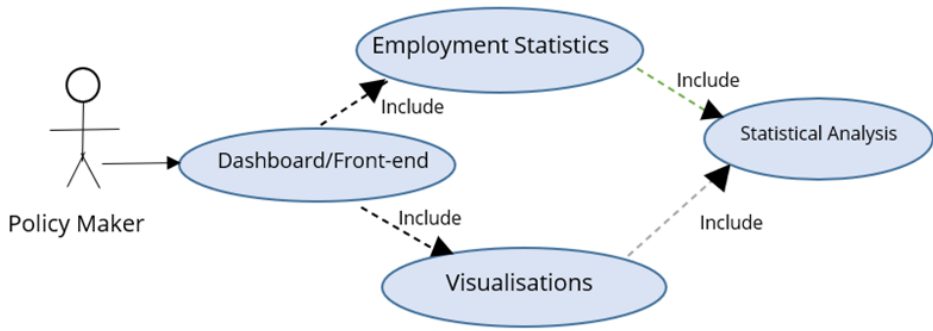
Section	Description
ID	SCE- PAUNRI -02
Title	Calculation capabilities
Description	Linked to scenario B; Process data and be able to perform calculations such as Sum: Adding the total number for a specified column. Count: Count the number of instances within a given column.
Actors	End User, Policy makers
Objectives	<ul style="list-style-type: none"> Count number of citizens Group age grounds Show total number of citizens claiming in a specific time period Support trend analysis
Pre-Conditions	Statistics related to data available from the front end of the PDT
Process Description	<ul style="list-style-type: none"> London borough of Camden uploads collected data PolicyCLOUD technological providers analyse the data Policy makers get results from scenario analysis London borough benefits from improved policy making process
Variations	N/A
Post-Conditions	Objectives completed
UML User Case Diagram	 <pre> graph LR PM[Policy Maker] --> DF[Dashboard/Front-end] DF -.-> Include ES[Employment Statistics] DF -.-> Include V[Visualisations] ES -.-> Include SA[Statistical Analysis] V -.-> Include SA </pre>

TABLE 57 - SCENARIO SCE-PAUNRI-01

2.4.3 Stakeholder Requirements

The following tables contain the initial list of the stakeholder requirements for the scenarios of this use case that were described in the previous subsection.

Section	Description
ID	REQ- PAUNRI-01
Title	Analysis capabilities
Level of detail	Stakeholder
Type	ENV
Description	<ul style="list-style-type: none"> Describe the nature of the data to be analysed. Explore the relation of the data to the underlying population. Create a model to summarize an understanding of how the data relates to the underlying population. Prove (or disprove) the validity of the model. Employ predictive analytics to run scenarios that will help guide future actions.
Additional Information	N/A
Actor	End User, Policy makers
Priority	MAN
Reference Use Case	SCE- PAUNRI -01, SCE- PAUNRI -02
Success Criteria	The success of this requirement can be tracked based on the feedback from the users that have completed questionnaires. The questionnaire will include clear questions that will indicate whether the user experience with the platform was positive or negative.
Expected delivery date	Before the end of M24

TABLE 58 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-01

Section	Description
ID	REQ- PAUNRI-02
Title	User interface
Level of detail	Stakeholder
Type	L&F
Description	The User interface should be straight forward, and pages and function buttons should load in a decent amount of time. Pages should not be overloaded with text and page layout should main a consistent standard.
Additional Information	N/A
Actor	The list of different actors that are related to this requirement
Priority	DES

Reference Use Case	SCE- PAUNRI -01, SCE- PAUNRI -02
Success Criteria	The success of this requirement can be tracked based from the feedback from users that have completed questionnaires. The questionnaire will include clear questions that will indicate whether the user experience with the platform was positive or negative.
Expected delivery date	Before the end of M24

TABLE 59 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-02

Section	Description
ID	REQ- PAUNRI-03
Title	Secure infrastructure
Level of detail	Stakeholder
Type	ENV
Description	The User interface should be straight forward, and pages and function buttons should load in a decent amount of time. Web sessions should timeout after a long period of inactivity.
Additional Information	N/A
Actor	End User, Policy makers
Priority	DES
Reference Use Case	SCE- PAUNRI -01, SCE- PAUNRI -02
Success Criteria	User can login successfully via secure connections that will throw time out errors after a long period of inactivity.
Expected delivery date	Before the end of M24

TABLE 60 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-03

Section	Description
ID	REQ- PAUNRI-04
Title	Help page/ Support documentation
Level of detail	Stakeholder
Type	SUP
Description	Help and support should be available for users if they ever become lost or unable to proceed due to lack of information. Support should be in the form of either a tool tip button or a hover cursor or separate support page.
Additional Information	N/A
Actor	End User, Policy makers
Priority	ENH

Reference Use Case	SCE- PAUNRI -01, SCE- PAUNRI -02
Success Criteria	User should be able to access documentation or a user help page that will assist and display the explanation for each of the toolkits action buttons
Expected delivery date	Before the end of M24

TABLE 61 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-04

Section	Description
ID	REQ- PAUNRI-05
Title	Visualisation capabilities
Level of detail	Stakeholder
Type	ENV
Description	The PolicyCLOUD toolkit should be able to produce some form of visualisations (e.g., line, pie and bar graphs) that can show assist with statistical/trend analysis. The figures and graph labels should be legible and displayed in English text.
Additional Information	Guidance from the example visualisations provided for each use case scenario is available upon request
Actor	End User, Policy makers
Priority	MAN
Reference Use Case	SCE- PAUNRI -01, SCE- PAUNRI -02
Success Criteria	Various types of graphs based on the dataset should be available to view on the front end of the PDT
Expected delivery date	Before the end of M24

TABLE 62 - STAKEHOLDER REQUIREMENT REQ - PAUNRI-05

3 Use case datasets and data regulatory constraints

This section contains the definition of all the available datasets that will be used in the scope of the PolicyCLOUD project, along with potential data regulatory constraints that might be needed to be enforced when these datasets are being accessed by third parties or are being collected and stored in a cloud environment, such as the deployment of the PolicyCLOUD platform outside of the proprietary's premises. The list of these requirements for data management and regulatory constraints will mainly drive the implementation of the corresponding components of the overall PolicyCLOUD architecture, mainly the data repository, the data fusion and the definition of the data governance model and the protection and privacy enforcement. These requirements are listed in the following two subsections and are presented per use case. An in-depth analysis regarding the ethical, legal, regulatory, and societal requirements related to datasets has been performed in D3.3 PolicyCLOUD's Societal and Ethical Requirements & Guidelines - M10 and updated in the context of D3.6 PolicyCLOUD's Societal and Ethical Requirements & Guidelines - M22.

3.1 Dataset Specifications

3.1.1 UC#1 - Participatory policies against radicalization

Section	Description
ID	DS-PPR-01
Title	Twitter
Description	Relevant posts published by users
Owner	Twitter
Licence/Privacy	Twitter license
Data type	Text and images
Type of Process (Stream or Static data)	streaming
Data Format	JSON
Data Store	N/A
Recommended API	REST API
Data Volume	N/A
Data Velocity	Every minute
Documentation	Twitter documentation

TABLE 63 - DATASET REQUIREMENT DS -PPR-01

Section	Description
ID	DS-PPR-02
Title	Facebook
Description	Relevant posts and comments published in open groups
Owner	Facebook
Licence/Privacy	Facebook license
Data type	Text and images
Type of Process (Stream or Static data)	streaming
Data Format	JSON
Data Store	N/A
Recommended API	REST API
Data Volume	N/A
Data Velocity	Every minute
Documentation	Twitter documentation

TABLE 64 - DATASET REQUIREMENT DS-PPR-02

Section	Description
ID	DS-PPR-03
Title	Reddit
Description	Relevant posts published by users
Owner	N/A
Licence/Privacy	N/A
Data type	Text, images
Type of Process (Stream or Static data)	Streaming
Data Format	JSON
Data Store	N/A
Recommended API	REST API
Data Volume	N/A
Data Velocity	Every minute
Documentation	Reddit documentation

TABLE 65 - DATASET REQUIREMENT FOR DS-PPR-03

Section	Description
ID	DS-PPR-04
Title	RSS Feeds
Description	Relevant news available on the web
Owner	N/A
Licence/Privacy	N/A
Data type	unstructured (i.e. text, article, image)
Type of Process (Stream or Static data)	stored on data repository
Data Format	HTML
Data Store	N/A
Recommended API	REST API
Data Volume	N/A
Data Velocity	Every hour
Documentation	

TABLE 66 - DATASET REQUIREMENT FOR DS-PPR-04

Section	Description
ID	DS-PPR-05
Title	Global Terrorism Database ⁵
Description	Open-source database including information on domestic and international terrorist attacks around the world from 1970 and includes more than 190,000 cases.
Owner	The National Consortium for the Study of Terrorism and Responses to Terrorism (START)
Licence/Privacy	Open source for research purposes, Licenses for commercial purposes
Data type	Structured (i.e., text, article, image)
Type of Process (Stream or Static data)	stored on data repository
Data Format	Structured data (.xlsx)
Data Store	N/A
Recommended API	REST API
Data Volume	Couple of GBs

⁵ <https://www.start.umd.edu/gtd/>

Data Velocity	Every week
Documentation	https://www.start.umd.edu/gtd/

TABLE 67 - DATASET REQUIREMENT FOR DS-PPR-05

Section	Description
ID	DS-PPR-06
Title	RAND Database of Worldwide Terrorism Incidents ⁶
Description	Open-source database including information of international and domestic terrorism incidents from 1968 through 2009, and now includes more than 40,000 cases.
Owner	RAND Corporation – National Security Research Division
Licence/Privacy	Open source for research purposes, Licenses for commercial purposes
Data type	Structured (i.e., text, article, image)
Type of Process (Stream or Static data)	Stored on data repository
Data Format	Structured data (.xlsx)
Data Store	N/A
Recommended API	REST API
Data Volume	Couple of GBs
Data Velocity	Every week
Documentation	https://www.rand.org/nsrd/projects/terrorism-incidents.html

TABLE 69 - DATASET REQUIREMENT FOR DS-PPR-05

⁶ <https://www.rand.org/nsrd/projects/terrorism-incidents.html>

3.1.2 UC#2 - Intelligent policies for the development of agrifood industry

In the following tables, the datasets of the use case *Intelligent policies for the development of agrifood industry* are described.

Section	Description
ID	DS-IIPDD-01
Title	CAP
Description	The Common Agricultural Policy (CAP) is the agricultural policy of the European Union. It implements a system of agricultural subsidies and other programmes. It was introduced in 1962 and has undergone several changes since then to reduce the cost (from 73% of the EU budget in 1985 to 37% in 2017) and to also consider rural development in its aims. It has been criticised on the grounds of its cost, and its environmental and humanitarian impacts.
Owner	Open Data Aragon (https://opendata.aragon.es/)
Licence/Privacy	It can be imported in PolicyCLOUD
Data type	Semi-structural
Type of Process (Stream or Static data)	Database
Data Format	Virtuoso (triplets), JSON or XML
Data Store	Virtuoso
Recommended API	Sparql
Data Volume	Millions but we require few of them
Data Velocity	Periodically after a couple of days
Documentation	https://opendata.aragon.es/datos/catalogo?texto=pac

TABLE 68 - DATASET REQUIREMENT FOR DS-IIPDD-01

Section	Description
ID	DS-IIPDD-02
Title	Wine register
Description	Sigpac reference, variety, cultivation year, area (hec)
Owner	Aragon Government
Licence/Privacy	It can be imported in PolicyCLOUD
Data type	Structural
Type of Process (Stream or Static data)	static

Data Format	csv
Data Store	Web links to files
Recommended API	REST API
Data Volume	1 Gbytes
Data Velocity	Periodically after a couple of days
Documentation	https://www.aragon.es/en/-/consultas-sigpac

TABLE 69 - DATASET REQUIREMENT FOR DS-IIPDD-02

Section	Description
ID	DS-IIPDD-03
Title	Production data
Description	Data production per grape variety
Owner	SARGA
Licence/Privacy	
Data type	Structural
Type of Process (Stream or Static data)	Static data
Data Format	Tables ODBC
Data Store	Microsoft SQL
Recommended API	JDBC
Data Volume	2 GBytes
Data Velocity	Periodically after a couple of days
Documentation	

TABLE 70 - DATASET REQUIREMENT FOR DS-IIPDD-03

Section	Description
ID	DS-IIPDD-04
Title	Twitter data
Description	Information provided by users about wine varieties and brands
Owner	N/A
Licence/Privacy	Twitter license
Data type	Text and images
Type of Process (Stream or Static data)	streaming
Data Format	JSON
Data Store	N/A

Recommended API	REST API
Data Volume	1 Terabyte for this scenario
Data Velocity	Every hour 10 minutes
Documentation	Twitter documentation

TABLE 71 - DATASET REQUIREMENT FOR DS-IIPDD-04

Section	Description
ID	DS-IIPDD-05
Title	Facebook data
Description	Information provided by users about wine varieties and brands
Owner	N/A
Licence/Privacy	Facebook license to be discussed
Data type	Text, images
Type of Process (Stream or Static data)	Streaming
Data Format	JSON
Data Store	N/A
Recommended API	REST API
Data Volume	1 Terabyte
Data Velocity	Every 10 minutes
Documentation	Facebook documentation

TABLE 72 - DATASET REQUIREMENT FOR DS-IIPDD-05

Section	Description
ID	DS-IIPDD-06
Title	Instagram
Description	Information provided by users about wine varieties and brands
Owner	N/A
Licence/Privacy	Instagram license
Data type	Unstructured (i.e., text, article, image)
Type of Process (Stream or Static data)	Streaming and stored on data repository
Data Format	JSON
Data Store	N/A
Recommended API	REST API
Data Volume	1 Terabyte

Data Velocity	Every 10 minutes
Documentation	Instagram

TABLE 73 - DATASET REQUIREMENT FOR DS-IIPDD-06

Section	Description
ID	DS-IIPDD-08
Title	News webpages/blogs
Description	News about wine, brands, production of wine
Owner	Newspapers
Licence/Privacy	N/A
Data type	Unstructured (i.e., text, article, image)
Type of Process (Stream or Static data)	Stored on data repository
Data Format	HTML
Data Store	N/A
Recommended API	REST API
Data Volume	Few gigabytes
Data Velocity	Every hour
Documentation	N/A

TABLE 74 - DATASET REQUIREMENT FOR DS-IIPDD-08

Section	Description
ID	DS-IIPDD-09
Title	E-commerce websites for wine use case
Description	Relevant e-commerce websites about wine varieties, brands, captured and processed for subsequent analysis in UC#2
Owner	Newspapers
Licence/Privacy	N/A
Data type	Unstructured (i.e., text, article, image)
Type of Process (Stream or Static data)	Stored on data repository
Data Format	HTML
Data Store	N/A
Recommended API	REST API
Data Volume	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store
Data Velocity	Every hour

Documentation	N/A
----------------------	-----

TABLE 75 - DATASET REQUIREMENT FOR DS-IIPDD-09

Section	Description
ID	DS-IIPDD-10
Title	Wine varieties and brands prices from the e-commerce
Description	Relevant e-commerce websites about wine varieties, prices, brands, captured and processed for subsequent analysis in UC#2
Owner	Newspapers
Licence/Privacy	N/A
Data type	Unstructured (i.e., text, article, image)
Type of Process (Stream or Static data)	Stored on data repository
Data Format	HTML
Data Store	N/A
Recommended API	REST API
Data Volume	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store
Data Velocity	Every hour
Documentation	N/A

TABLE 76 - DATASET REQUIREMENT FOR DS-IIPDD-10

3.1.3 UC#3 - Facilitating urban policy making and monitoring through crowdsourcing data analysis

In the following tables, the datasets of the use case *Facilitating urban policy making and monitoring through crowdsourcing data analysis* are described.

Section	Description
ID	DS-RLIMP-01
Title	SofiaMunicipalitySignals
Description	Signal from citizens. Citizens file a signal in the Contact Centre in the categories concerned – transport, parking, road infrastructure, waste collection and waste disposal and air quality. They choose a specific type of problem for each category. In a free text they describe what is the problem. They also specify the geographical location and the exact address of the signal.
Owner	Sofia Municipality
Licence/Privacy	The data are legible to be imported in PolicyCLOUD environment
Data type	Structured data
Type of Process (Stream or Static data)	The data is static. However, the status and resolution of signals changes and can be updated periodically. It can also be updated with additional records.
Data Format	XML or JSON
Data Store	MS SQL Server
Recommended API	Generating XML or JSON file with the necessary data directly or through REST API
Data Volume	Data per month 5000 * 70 KB/ month = 350 MB/ month However, this monthly data is estimated based on the signals coming through the web-based system only. In April 2020 Sofia Municipality is launching a mobile application in addition to the web-based system, so volume of signals is expected to increase
Data Velocity	5000 signals per month However, this monthly data is estimated based on the signals, coming through the web-based system only. In April 2020 Sofia Municipality is launching mobile app in addition to the web-based system, so volume of signals is expected to increase
Documentation	The data is not public. A limited number of signals (the latest 20) can be seen on https://call.sofia.bg/

TABLE 77 - DATASET REQUIREMENT FOR DS-RLIMP-01

Section	Description
ID	DS-RLIMP-02
Title	AirData
Description	Real time IoT sensors data for monitoring and measurement of the air quality. https://air.sofia.bg/

Section	Description
	Measurement of air conditions at given location. Types of the measurements are: <ul style="list-style-type: none"> • average concentration/per hour of fine dust particles - 10 micrometre, 2.5 micrometre, nitrogen dioxide (NO₂), ozone (O₃) and carbon oxide (CO) • average concentration/per day of Polycyclic Aromatic Hydrocarbon for 2015-2019 period
Owner	Executive Environment Agency
Licence/Privacy	The data can be imported in PolicyCLOUD environment
Data type	Structured data
Type of Process (Stream or Static data)	The data is static
Data Format	XML or JSON
Data Store	MS SQL Server
Recommended API	Generating XML or JSON file with the necessary data directly or through REST API
Data Volume	7 stations
Data Velocity	7 stations
Documentation	air.sofia.bg

TABLE 78 - DATASET REQUIREMENT FOR DS-RLIMP-02

3.1.4 UC#4 - Predictive analysis towards unemployment risks identification and policy making

In the following tables, the datasets of the use case are described.

Section	Description
ID	DS-PAUNRI-01
Title	Unemployment Claimant Count LATEST
Description	The unemployment claimant count dataset includes details of the amount of people claiming benefits related to Jobseeker's Allowance (JSA). JSA is a government funded benefit to support unemployed citizens looking for work. The data is provided to Camden from the Office of National Statistics.
Owner	Open and publicly available dataset provided via the Office for National Statistics
Licence/Privacy	Due to the data being open this means that it is available to the public. However, any form of commercial usage would have to be reviewed as this may be prohibited. Given the fact that the owner is notified, the dataset can be imported into a PolicyCLOUD environment

Section	Description
	Licence: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
Data type	Structural (CSV/JSON) both available)
Type of Process (Stream or Static data)	This dataset will be stream data that will be periodically updated on a monthly basis
Data Format	The data can be stored as either CSV or JSON
Data Store	TYLER TECH open data platform
Recommended API	API (Returns 1000 records by default see API documentation) https://opendata.camden.gov.uk/People-Places/JSA-And-UC-Claimants-In-Camden/x2rm-3zds/data
Data Volume	2000(Records)/500KB
Data Velocity	One monthly update of 2000records equivalent of 500KB
Documentation	API - https://opendata.camden.gov.uk/resource/x2rm-3zds.json

TABLE 79 - DATASET REQUIREMENT FOR DS-PAUNRI-01

4 Platform Roles

The following table contains the list of all the different roles of actors that are related with the development, deployment, operation and usage of all solutions that are offered by the PolicyCLOUD platform, along with a description of these roles.

ID	Name	Description
ROL-01	Data Owners	PolicyCLOUD provides a cloud Gateway component that can be used in order to push data coming from different data owners in various formats that can be either static or streaming data. This data is finally stored in the data store of the platform and is accessible by its analytical tools.
ROL-02	Data Engineers	PolicyCLOUD offers via the Data Analytics component a framework to register their analytical tools that can use the common data repository of the project in order to perform analytical tasks on the stored data. It can also rely on the intermediate results of other tools and feed with them his/her model.
ROL-03	Policy Makers	PolicyCLOUD offers the Policy Development Toolkit that allows the policy makers to create and evaluate new policies in different domains, associate the policies with specific KPIs and validate them by triggering the execution of one of the PolicyCLOUD's analytical tools by seeing the results visualized in the toolkit.
ROL-04	Data Scientists	PolicyCLOUD offers to the data scientists the Data Marketplace in order to explore and validate the provided analytical tools with different target datasets and experimenting with extended datasets.

TABLE 80 - PLATFORM ROLES

5 System Requirements

This section will present the system requirements of PolicyCLOUD, which represent the technical specifications for the platform at the systemic level. They define the services along with the functionalities and their interfaces of the major building blocks that formulate the overall platform and need to meet the stakeholder requirements. They answer to what characteristics the system needs to possess and to what degree to satisfy the stakeholders' requirements.

The system requirements that are defined in this Section are organized per major building block that corresponds to a specific capability. All these capabilities are depicted in the full stack that is presented in Figure 2. As it can be depicted, PolicyCLOUD platform proves a full stack of capabilities that aim for the data acquisition and collection via its gateways, the persistent store of data in the data repository, the deployment of numerous and heterogeneous analytical tools that can make use of this data, the incorporation of reusable models exploited by these tools, and finally, the creation and evaluation of policies via the relevant policy development toolkit, that relies on the use of the models and tools of the underlying layer, and the experimentation with additional datasets that can be discovered and exploitable via the use of the Marketplace.

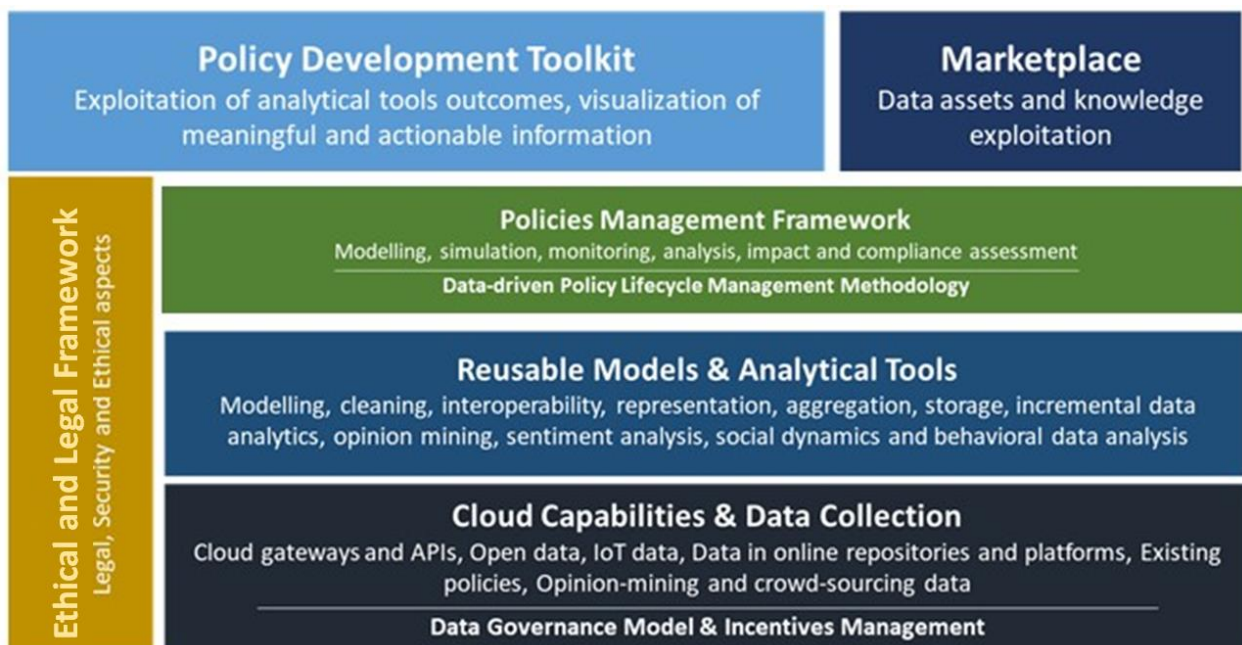


FIGURE 2 - POLICYCLOUD CORE SYSTEM CAPABILITIES

In more detail, PolicyCLOUD consists of the following different capability categories:

- **Cloud Capabilities and Data Collection** - This layer consists of all the operations that are required for cloud registration and resource provisioning in order for all the platform components and analytical tools to be instantiated and deployed. It additionally consists of various mechanisms for data acquisition, collection, cleaning and persistent storage, while also providing the means

for a seamless manner for data retrieval of data that might exist inside the platform, or that can be accessed in cases a dataset must stay on-premise and cannot be loaded to the platform.

- **Data Governance Model and Incentives Management** - This layer is taking care of the various data regulatory constraints that are imposed by the use cases and the owners of the data, in order to allow them access to specific data resources or not. It is used by the data collection mechanisms in order to allow for the data acquisition, while at the same time it is used by the seamless data retrieval component of the polyglot data store in order to allow the establishment of a data connection to external resources for data retrieval.
- **Reusable Models and Analytical Tools** - This layer contains all the various analytical tools that are natively supported and implemented by the PolicyCLOUD platform, such as opinion mining and sentiment analysis, behavioural data analytics, situational knowledge acquisition, social dynamics, etc. It also contains the reusable models that can be shared by all tools, thus increasing the level of interoperability of all the deployed tools.
- **Policy Development Toolkit** - This layer offers a framework for policy makers and domain experts to define implement and evaluate a new policy that is related with specific KPIs and trigger the execution of an analytical task in order to validate the results against the already defined KPIs. It has access to the underlying analytical tools provided by the underlying layer and can make use of intermediate or previous produced results to validate the KPIs in the process of the time, making use of the data stored persistently in the PolicyCLOUD or data that can be accessed via the use of the polyglot capabilities of the data repository of the platform.
- **Data Marketplace** - This layer of PolicyCLOUD provides to the data analysts and policy experts the ability to discover newly available datasets that have been stored into the platform by various data providers, and either trigger the data acquisition process in order to ingest the corresponding datasets into the data repository, so that they can be available by the analytical tools, or to provide the capability to connect remotely to the external data source, if this is allowed by the data regulatory constraints that are validated by the Data Governance Model.
- **Legal and Ethical Framework** - A layer that included a framework that consists of legal, regulatory, ethical, and societal requirements that will drive the project during its development phase but also throughout its exploitation. General requirements and use case specific requirements have been provided in the D3.3 and D3.6 then will be updated at M36.
- **Policy Management Framework** - The Policies Management Framework incorporates services for the identification of the required KPIs in order to model the policies and identify potential interdependencies with other policies within and across sectors at different levels (e.g. local, national, etc). The framework also includes tools for collecting evidence monitoring information both from the engaged citizens and from the population targeted by the policies, while also assessing the compliance to these policies and thus assessing the policies impact (based on the identified KPIs). More information can be found in "Cross-sector Policy Lifecycle Management: Design and Open Specification 2", D5.4 which has been submitted in M20 and in "Cross-sector Policy Lifecycle Management: Software Prototype 2, D5.5 which has been submitted in M22.

The following subsections contain the list of the system requirements that are imposed by each of the aforementioned system capabilities of the PolicyCLOUD platform.

5.1 Cloud Capabilities and Data Collection

Section	Description
ID	REQ- SY-CCDC-01
Title	Cloud data storage should scale out in order to store data whose size is getting increased
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	In cases that a use case is increasing the volume of the data to be stored, the cloud data storage should be able to scale out so that all data can be persistently stored
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02
Source	N/A
Success Criteria	Stress the storage to reach its limits and check that it does not fail
Expected delivery date	M36 (3rd version of prototypes)

TABLE 81 - SYSTEM REQUIREMENT REQ -SY-CCDC-01

Section	Description
ID	REQ- SY-CCDC-02
Title	Provide adequate compute, memory and storage to fit the needs of the project
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	Considering the pledged resources allocated by the provider (see Section 6.1), the testbed is composed by 308 GB of RAM/64 CPU cores= 4.8GB of RAM per core, 2TB of block storage and 50-100MB of object storage
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02
Source	N/A
Success Criteria	All components of PolicyCLOUD can be deployed with the available resources
Expected delivery date	M36 (3rd version of prototypes)

TABLE 82 - SYSTEM REQUIREMENT REQ-SY-CCDC-02

5.2 Data Governance Model and Incentives Management

Section	Description
ID	REQ- SY-DGMIM-01
Title	Incentive Management Visualizations
Level of detail	System
Type	FUNC (function)
Description	The Incentives Management component(s) will interact with the Policy Maker and with the participants through the PDT
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-03
Source	PDT Technology
Success Criteria	N/A
Expected delivery date	M24 (2nd version of prototypes)

TABLE 83 - SYSTEM REQUIREMENT REQ - SY-DGMIM-01

Section	Description
ID	REQ- SY-DGMIM-02
Title	Incentive Management REST API
Level of detail	System
Type	FUNC (function)
Description	The component(s) provided in the context of the Incentives Management will expose their features through an interface in the form of a REST API
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	REST technology
Success Criteria	N/A
Expected delivery date	M12 (1st version of prototypes)

TABLE 84 - SYSTEM REQUIREMENT REQ- SY-DGMIM-02

Section	Description
ID	REQ- SY-DGMIM-03
Title	Incentive Management Storage Backend
Level of detail	System
Type	DATA (data)
Description	The domain objects resulting from the Incentives Management component(s) will be stored in the PolicyCLOUD storage - Users, Incentives, Task
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	PolicyCLOUD data repository
Success Criteria	N/A
Expected delivery date	M12 (1st version of prototypes)

TABLE 85 - SYSTEM REQUIREMENT REQ- SY-DGMIM-03

Section	Description
ID	REQ- SY-DGMIM-04
Title	Incentive Management component(s) Reusability
Level of detail	System
Type	USE (Usability Requirements), SUP (Maintainability and Support Requirements)
Description	Existing open source solution will be evaluated before developing new Incentives Management component(s)
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	N/A
Expected delivery date	M12 (1st version of prototypes)

TABLE 89 - SYSTEM REQUIREMENT REQ- SY-DGMIM-04

Section	Description
---------	-------------

ID	REQ- SY-DGMIM-05
Title	Self-hosting incentive management tool
Level of detail	System
Type	ENV (Operational/Environment Requirements), SUP (Maintainability and Support Requirements)
Description	In case the incentive management features will be accomplished by an existing open source solution, the tool might be self-hosted by the PolicyCLOUD infrastructure
Additional Information	N/A
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	N/A
Expected delivery date	M24 (2nd version of prototypes)

TABLE 86 - SYSTEM REQUIREMENT REQ- SY-DGMIM-05

Section	Description
ID	REQ- SY-DGMIM-06
Title	Request authentication before authorization when accessing data
Level of detail	System
Type	FUNC (function)
Description	To securely allow data to be accessed, a user must be authenticated through an Identity Provider, and then the request can be authorized by the ABAC Engine that enforces data governance policies
Additional Information	The identity provider can also be seen as part of the platform integration; nevertheless, this requirement is added here as Identity Provider is needed for the Privacy Enforcement Mechanism
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	Allow access to data based through the usage of authenticated user
Expected delivery date	M12 (1st version of prototypes)

TABLE 87 - SYSTEM REQUIREMENT REQ- SY-DGMIM-06

Section	Description
---------	-------------

ID	REQ- SY-DGMIM-07
Title	Controlled access to the identity provider and data to specific IPs and domains
Level of detail	System
Type	FUNC (function)
Description	Requests allowed by the ABAC and the identity provider should be able to be limited to specific URLs provided by the pilots
Additional Information	The identity provider can also be seen as part of the platform integration; nevertheless, this requirement is added here as Identity Provider is needed for the Privacy Enforcement Mechanism
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	Limit access to the Identity Provide to specific domains. Limit access to data based on request IP
Expected delivery date	M24 (2nd version of prototypes)

TABLE 88 - SYSTEM REQUIREMENT REQ- SY-DGMIM-07

Section	Description
ID	REQ- SY-DGMIM-08
Title	Separation and isolation of users based on use case need
Level of detail	System
Type	ENV (Operational/Environment Requirements), SUP (Maintainability and Support Requirements)
Description	A separate and isolated Identity Provider client will serve each pilot for each use case
Additional Information	The identity provider can also be seen as part of the platform integration; nevertheless, this requirement is added here as Identity Provider is needed for the Privacy Enforcement Mechanism
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	Isolation of use case clients
Expected delivery date	M24 (2nd version of prototypes)

TABLE 89 - SYSTEM REQUIREMENT REQ- SY-DGMIM-08

Section	Description
---------	-------------

ID	REQ- SY-DGMIM-09
Title	User attributes for ABAC Engine evaluation
Level of detail	System
Type	ENV (Operational/Environment Requirements), SUP (Maintainability and Support Requirements)
Description	All the required user attributes for the ABAC authentication and authorization should be requested during the user's registration
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	User attributes can be added during the registration phase
Expected delivery date	M12 (1st version of prototypes)

TABLE 90 - SYSTEM REQUIREMENT REQ- SY-DGMIM-09

Section	Description
ID	REQ- SY-DGMIM-10
Title	Updating user attributes for ABAC Engine evaluation
Level of detail	System
Type	ENV (Operational/Environment Requirements), SUP (Maintainability and Support Requirements)
Description	User attributes should be able to be updated, either by the user or by the platform administrator
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	User attributes are able to be updated securely
Expected delivery date	M24 (2nd version of prototypes)

TABLE 91 - SYSTEM REQUIREMENT REQ- SY-DGMIM-10

Section	Description
ID	REQ- SY-DGMIM-11

Title	Data Governance should be extensive in order to support access based on attributes relevant to use case needs
Level of detail	System
Type	ENV (Operational/Environment Requirements), SUP (Maintainability and Support Requirements)
Description	The attributes required by the use cases shall be collected, and the Data Governance Model and Privacy Enforcement Mechanism will be able to provide access control based on them. These attributes will be related to the user, to the object to be accessed and other environmental attributes, based on the ABAC model
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	Data Governance Model and Privacy Enforcement Mechanism
Success Criteria	Data Governance Model and Privacy Enforcement Mechanism support the ABAC model
Expected delivery date	M24 (2nd version of prototypes)

TABLE 92 - SYSTEM REQUIREMENT REQ- SY-DGMIM-11

5.3 Reusable Models and Analytical Tools

This section provides the requirements for the Data Acquisition and Analytics (DAA) layer which is the main outcome of WP4 – “Reusable Models and Analytical Tools”.

Section	Description
ID	REQ- SY-RMAT-01
Title	Minimum hardware requirements
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	The analytical components require minimum 8 CPU cores and 16G of memory to operate
Additional Information	At least for sentiment analysis, opinion mining and situational knowledge analysis
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The analytical components should be running with no problems

Expected delivery date	M12 (1st version of prototypes)
-------------------------------	---------------------------------

TABLE 93 - SYSTEM REQUIREMENT REQ -SY-RMAT-01

Section	Description
ID	REQ- SY-RMAT-02
Title	Define a meta-schema for schema definitions to be used for data sources registrations and resulted analytics
Level of detail	System
Type	DATA (data)
Description	A data schema will be defined to be able to process the data coming from external data sources and from the PolicyCLOUD datastore
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	Data can be correctly processed
Expected delivery date	M12 (1st version of prototypes)

TABLE 94 - SYSTEM REQUIREMENT REQ -SY-RMAT-02

Section	Description
ID	REQ- SY-RMAT-03
Title	Define an interface to be used by the Policy layer (PDT)
Level of detail	System
Type	DATA (data)
Description	An interface will be developed by defining a set of parameters that should be included when this component is executed by the PDT
Additional Information	Those parameters are related to the input parameters that the component requires, the type of data source that is being analysed, the type of output required to be visualized, and more as might be found at later phase
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	The interface satisfied all required combinations of applying analytics by the PDT
Expected delivery date	M12 (1st version of prototypes)

TABLE 99 - SYSTEM REQUIREMENT REQ-SY-RMAT-03

Section	Description
ID	REQ- SY-RMAT-04
Title	Register all analytic tools as OpenWhisk actions
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	DAA layer would support analytic tools written in Python and Java/Scala and provide Openwhisk implementation guidelines for data engineers. For some analytic tools the built-in supported Docker images in OpenWhisk can be used, while for others it will be required to create a special Docker image. Docker images should provide and specify all required packages. Type of analytic tool used should be specified during analytic registration at DAA and artifacts should be put in Policy Cloud repository.
Additional Information	This is according to the serverless-based architecture of the DAA layer
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	All analytical tools of the DAA layer can be activated in a serverless manner as OpenWhisk actions
Expected delivery date	M24 (2nd version of prototypes)

TABLE 95 - SYSTEM REQUIREMENT REQ-SY-RMAT-04

Section	Description
ID	REQ- SY-RMAT-05
Title	Stream data analysis
Level of detail	System
Type	DATA (data)
Description	The policy maker shall be able to perform analysis (at least opinion mining, sentiment and situational analysis) over continuous data coming from streaming data sources (i.e., social networks)
Additional Information	Their feasibility will depend on the readiness of the acquisition (cloud gateways) and pre-processing (data cleaning, data fusion and data interoperability) components.
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Demonstrated end-to-end use cases involving streaming analytics

Expected delivery date	M24 (2nd version of prototypes)
-------------------------------	---------------------------------

TABLE 96 - SYSTEM REQUIREMENT REQ-SY-RMAT-05

Section	Description
ID	REQ- SY-RMAT-06
Title	Kafka streams messaging
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	Continues data coming from streaming data sources such as Twitter Streaming API shall be published on a Kafka cluster
Additional Information	At least for sentiment analysis, opinion mining and situational knowledge analysis
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	Demonstrated end-to-end use cases involving streaming analytics through Kafka
Expected delivery date	M12 (1st version of prototypes)

TABLE 97 - SYSTEM REQUIREMENT REQ-SY-RMAT-06

Section	Description
ID	REQ- SY-RMAT-07
Title	Batch data analysis
Level of detail	System
Type	DATA (data)
Description	The policy maker shall be able to perform analysis (at least opinion mining and sentiment analysis) over collection of data persisted in the PolicyCLOUD storage
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Demonstrated end-to-end use cases involving batch data analytics
Expected delivery date	M12 (1st version of prototypes)

TABLE 98 - SYSTEM REQUIREMENT REQ-SY-RMAT-07

Section	Description
ID	REQ- SY-RMAT-08
Title	Standard API for database accessing
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	All datastores managed in PolicyCLOUD must provide a standard and common API for the data access/manipulation by the analytical components
Additional Information	At least for sentiment analysis, opinion mining and situational knowledge analysis
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The analytical components should be running with no problems
Expected delivery date	M24 (2nd version of prototypes)

TABLE 99 - SYSTEM REQUIREMENT REQ-SY-RMAT-08

Section	Description
ID	REQ- SY-RMAT-09
Title	External database analysis
Level of detail	System
Type	DATA (data)
Description	The policy maker shall be able to perform analysis over external databases
Additional Information	N/A
Priority	ENH (possible future enhancement)
Reference Use Case	UC#1, UC#2
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Demonstrated end-to-end use cases involving external data sources analytics
Expected delivery date	M36 (3rd version of prototypes)

TABLE 100 - SYSTEM REQUIREMENT REQ-SY-RMAT-09

Section	Description
ID	REQ- SY-RMAT-10
Title	Standard API for logging
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	Provide API for PolicyCLOUD central logging
Additional Information	Based on cloud compatibly logging infrastructure such as Logstash
Priority	ENH (possible future enhancement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	All WP4 components can append log entries to the central logging mechanism
Expected delivery date	M36 (3nd version of prototypes)

TABLE 101 - SYSTEM REQUIEREMENT REQ-SY-RMAT-10

5.4 Policy Management Framework

Section	Description
ID	REQ-SY-PMF-01
Title	Policy Model Editing
Level of detail	System
Type	FUNC (function), DATA (data), L&F (look & feel), USE (usability)
Description	The Policy maker – via PDT – should be able to view the structure of an existing policy model, modify and save the policy model (if she/he has the proper rights / ownership). Otherwise, he/she has access only to view-only functions.
Additional Information	Creation of new Policy Models following system templates, or by copying existing ones
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	Policy Models Editing capability depending on user rights
Expected delivery date	M12 (View structure) / M24 (Editing) / M36 (Full version)

TABLE 102 - SYSTEM REQUIREMENT REQ-SY-PMF-01

Section	Description
ID	REQ-SY-PMF-02
Title	Integration with the PDT Backend
Level of detail	System
Type	FUNC (function)
Description	The Policy Model Editor should communicate with the PDT Backend to retrieve entities (stored KPI's, Domains, Stakeholders, etc.)
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	To persistently store the created policy to the PDT
Expected delivery date	M12 (1st version of the prototypes)

TABLE 103 - SYSTEM REQUIREMENT REQ-SY-PMF-02

Section	Description
ID	REQ-SY-PMF-03
Title	Integration into the PDT
Level of detail	System
Type	FUNC (function)
Description	The Policy Model Editor should be integrated into the PDT frontend user interface
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	To store the created policy to the PDT
Expected delivery date	M24 (2nd version of the prototypes)

TABLE 104 - SYSTEM REQUIREMENT REQ-SY-PMF-03

Section	Description
ID	REQ-SY-PMF -04
Title	User Help
Level of detail	System
Type	L&F (look & feel), USE (usability)
Description	The Policy Modeling Editor (PME) should support the policy maker with hints, action descriptions and guides as she/he performs policy creation/modification/verification actions
Additional Information	The users should also be supported during the selection of policies/ KPIs/ Analytics actions
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A policymaker can, with relative ease, explore PDT and perform policymaking tasks
Expected delivery date	M12 (hints) / M24 (descriptions) / M36 (guides)

TABLE 105 - SYSTEM REQUIREMENT REQ-SY-PMF -04

5.5 Policy Development Toolkit

Section	Description
ID	REQ- SY-PDT-01
Title	Analytical tools should expose a REST interface to allow their invocation from the PDT
Level of detail	System
Type	FUNC (function)
Description	The PDT must be able to invoke an analytical tool via a standard REST interface that is being registered into its catalogue
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	All tools must be accessible and be able to be invoked by the PDT
Expected delivery date	M12 (1st version of prototypes)

TABLE 106 - SYSTEM REQUIREMENT REQ - SY-PDT-01

Section	Description
ID	REQ- SY-PDT-02
Title	Analytical tools must register the parameters that are required as input
Level of detail	System
Type	FUNC (function)
Description	Each tool accepts a different type of parameters. These should be retrievable by the PDT in order to guide the user to fill those parameters and invoke the tool accordingly
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	Parameters must be serializable and follow a common format
Success Criteria	Analytical tool can be invoked with parameters of the same format
Expected delivery date	M12 (1st version of prototypes)

TABLE 107 - SYSTEM REQUIREMENT REQ - SY-PDT-02

Section	Description
ID	REQ- SY-PDT-03
Title	Visualizations of Analytics Results via PDT
Level of detail	System
Type	FUNC (function), L&F (look & feel)
Description	Policy makers should be able to view the results of the analytics via PDT
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	Policy makers can see the visualization of the results from the requested analytics
Expected delivery date	M36

TABLE 108 - SYSTEM REQUIREMENT REQ- SY-PDT-03

Section	Description
ID	REQ- SY-PDT-05
Title	User Authentication & Authorization
Level of detail	System
Type	FUNC (function), USE (usability)
Description	PDT User should be able to authenticate using her/his credentials into the system (Login). The content will vary depending on the credentials
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A user supplies the credentials and enters (logins) into the platform. Also, a user can edit her/his policy models, but e.g. only view other / system policy models
Expected delivery date	M12 (Authentication) / M24 (Authorization)

TABLE 109 - SYSTEM REQUIREMENT REQ- SY-PDT-05

Section	Description
ID	REQ- SY-PDT-06
Title	User Notifications on Analytics Progress

Section	Description
Level of detail	System
Type	FUNC (function), L&F (look & feel), USE (usability)
Description	PDT User should be informed about the status changes in the processing of Analytics Requests
Additional Information	The user should be informed if an analytics request has failed and the related reason. Also, a time estimation for the completion should be given. Finally, the user should be notified for the completion of the process
Priority	DES
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A policy maker who submitted an analysis request, is being notified as the request is being executed regarding its status and time expectations
Expected delivery date	M24 (Initial) / M36 (Full)

TABLE 110 - SYSTEM REQUIREMENT REQ- SY-PDT-06

Section	Description
ID	REQ- SY-PDT-07
Title	User Help
Level of detail	System
Type	L&F (look & feel), USE (usability)
Description	The PDT should support the policy maker with hints, action descriptions and guides as she/he performs policy creation/modification/verification actions
Additional Information	The users should also be supported during the selection of policies/ KPIs/ Analytics actions
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A policy maker can, with relative ease, explore PDT and perform policymaking tasks
Expected delivery date	M12 (hints) / M24 (descriptions) / M36 (guides)

TABLE 111 - SYSTEM REQUIREMENT REQ- SY-PDT-07

Section	Description
ID	REQ- SY-PDT-08
Title	Integration with the Data Analytics Layer

Level of detail	System
Type	FUNC (function)
Description	The PDT should communicate with the Data Analytic framework to retrieve the list of available tools and to send requests for analytics. It should also check the status of the pending jobs and retrieve the results from it
Additional Information	More information has been provided in the D4.3 deliverable
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	A policy maker can, with relative ease, explore PDT and perform policymaking tasks
Expected delivery date	M12 (1st version of the prototypes)

TABLE 112 - SYSTEM REQUIREMENT REQ- SY-PDT-08

Section	Description
ID	REQ- SY-PDT-09
Title	User Transaction History
Level of detail	System
Type	L&F (look & feel), USE (usability)
Description	The Policy Maker (User) should be able to see his/her past invocations of analytics tools along with their results, in relation with her policies/KPIs
Additional Information	Capability to the User to delete transactions (e.g., duplications in parameter, etc.)
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A policy maker can browse his/her transactions history and remove unneeded ones
Expected delivery date	M24 (2nd version of prototypes)

TABLE 113 - SYSTEM REQUIREMENT REQ- SY-PDT-09

Section	Description
ID	REQ- SY-PDT-10
Title	User Authentication & Authorization
Level of detail	System

Type	FUNC (function), USE (usability)
Description	Policy Maker should be able to authenticate using her/his credentials into the system (Login). The content will vary depending on the credentials
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	A User supplies the credentials and enters (logins) into the platform, and can compose a new policy model and submit in the PDT
Expected delivery date	M12 (Authentication) / M24 (Authorization)

TABLE 114 - SYSTEM REQUIREMENT REQ- SY-PDT-10

Section	Description
ID	REQ- SY-PDT-12
Title	Allow for integration with other tools for policy making
Level of detail	System
Type	FUNC (function)
Description	The PDT backend should be able to import policies defined in other tools (i.e., Politika) and be able to be exploited by the end-users of PolicyCLOUD
Additional Information	N/A
Priority	OPT (Optional)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	Policies defined in Politika can be retrieved by PDT
Expected delivery date	M36 (3rd version of the prototypes)

TABLE 115 - SYSTEM REQUIREMENT REQ- SY-PDT-12

5.6 Data Marketplace

Section	Description
ID	REQ-SY-DM-01
Title	User registration and authorization
Level of detail	System
Type	FUNC (function), USE (usability)
Description	The Data Marketplace requires registration and authorization in order to provide access to its contents. The users will be able to benefit from the Data Marketplace offerings and functionalities using their accounts
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	Successfully registered and logged in users to the Data Marketplace

Expected delivery date	M24 (2nd version of prototypes)
-------------------------------	---------------------------------

TABLE 116 - SYSTEM REQUIREMENT REQ- SY-DM-01

Section	Description
ID	REQ-SY-DM-02
Title	Compliance to regulations
Level of detail	System
Type	FUNC (function), DATA (data), USE (usability)
Description	The Data Marketplace should consider, comply, and inform the end-users on privacy issues (i.e., GDPR)
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	In order to comply to different regulations (e.g., GDPR), the Data Marketplace makes available the retrieval of users' data and let them know for their usage
Expected delivery date	M24 (2nd version of prototypes)

TABLE 117 - SYSTEM REQUIREMENT REQ- SY-DM-02

Section	Description
ID	REQ-SY-DM-03
Title	Assets' automated provision
Level of detail	System
Type	FUNC (function), DATA (data), USE (usability)
Description	The Data Marketplace has to implement functionalities in order to enable users to provide their assets by themselves
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	The users will provide their assets to the Data Marketplace, through appropriately formatted forms and interfaces
Expected delivery date	M24 (2nd version of prototypes)

TABLE 118 - SYSTEM REQUIREMENT REQ- SY-DM-03

Section	Description
ID	REQ-SY-DM-04
Title	Assets' management by the providers
Level of detail	System
Type	FUNC (function), USE (usability)
Description	The providers of the Data Marketplace's assets have to manage their offerings to the platform. In particular, they have to be able to perform all CRUD (Create, Read, Update and Delete) operations to both the assets and their descriptions
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	The Data Marketplace will implement functionalities to handle these operations and the users will be able to perform them by themselves
Expected delivery date	M24 (2nd version of prototypes)

TABLE 119 - SYSTEM REQUIREMENT REQ- SY-DM-04

Section	Description
ID	REQ-SY-DM-05
Title	Searching assets
Level of detail	System
Type	FUNC (function), USE (usability)
Description	The search functionality is a vital requirement for the most services in order to reduce the number of objects returned by a query. Thus, the users of the Data Marketplace have to be able to search for assets based on metadata that will be the contents of the descriptions
Additional Information	The descriptions are JSON formatted files that will contain metadata about the described assets
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	The Data Marketplace will implement functionalities that will enable the filtering of information and through them, the users will be able to retrieve assets related to their needs

Expected delivery date	M24 (2nd version of prototypes)
-------------------------------	---------------------------------

TABLE 120 - SYSTEM REQUIREMENT REQ- SY-DM-05

Section	Description
ID	REQ-SY-DM-06
Title	Definition of data schemas for users and assets
Level of detail	System
Type	FUNC (function), DATA (data), USE (usability)
Description	The Data Marketplace can handle data for its users and its offering assets. It has to have a standard schema for the collected data to present the information to the end-users. The information of the assets will be the content of the description files
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	N/A
Expected delivery date	M24 (2nd version of prototypes)

TABLE 121 - SYSTEM REQUIREMENT REQ- SY-DM-06

Section	Description
ID	REQ-SY-DM-07
Title	Implementation of back-end instance for assets' management
Level of detail	System
Type	FUNC (function), ENV (Operational/Environment)
Description	The Data Marketplace will consist of two components. The first is the back-end where the assets will be stored and managed by the implemented functionalities
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	REST API using Python's Flask framework and WSGIs Gunicorn and NGINX. MongoDB Database and GridFS specification will be used for assets' storage
Success Criteria	The implementation of the back-end and its functionalities
Expected delivery date	M24 (2nd version of prototypes)

TABLE 122 - SYSTEM REQUIREMENT REQ- SY-DM-07

Section	Description
ID	REQ-SY-DM-08
Title	Definition and implementation of front-end interfaces for assets' presentation
Level of detail	System
Type	FUNC (function), ENV (Operational/Environment), L&F (Look and Feel)
Description	The Data Marketplace will consist of two components. The second is the front-end, a web-based server that will present the contents of the back-end to the end-users, with a friendly User Interface
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	WordPress with various plugins, Web technologies (HTML, CSS, JavaScript, PHP, etc.)
Success Criteria	The implementation of the front-end and its interfaces
Expected delivery date	M24 (2nd version of prototypes)

TABLE 123 - SYSTEM REQUIREMENT REQ- SY-DM-08

5.7 Legal and Ethical Framework

To maximize societal acceptability and trust in PolicyCLOUD, and in the policies developed through it, the PolicyCLOUD consortium (the “**Consortium**”) is aware of the necessity of providing extensive and in-depth analyses of the legal, regulatory, societal, and ethical concerns which may be raised by PolicyCLOUD, its intended functions and features, and its potential use by current and future end-users. In particular, it is vital not only to ensure the lawfulness of the PolicyCLOUD solution, but also that it is designed in consideration of its potential ethical impact (at an individual and societal level), to ensure PolicyCLOUD’s acceptability as a solution to be leveraged by public authorities in their policy-making activities, as well as the trustworthiness of the platform (and of the output it generates) among public authorities, citizens and society at large. The goal of these analyses is to identify specific and practical requirements which can be optimally embedded into the design of the solution, supported by a thorough evaluation as to the extent to which this has been successful.

The Legal and Ethical Framework (the current iteration of which is reflected in D3.3 and in D3.6) identifies and structures these requirements, considering general requirements that can be considered as common and applicable to the PolicyCLOUD platform as a whole (in particular, issues arising from use of cloud-based infrastructure, big data, artificial intelligence, and personal data processing, among others), more specific requirements that refer to individual components of the platform (e.g., the data cleaning

tools, the analytical tools, the Policy Development Toolkit, the Policy Modelling Editor), and use-case specific analyses, considering the four current use cases.

In defining these requirements, particular attention is given to privacy and data protection laws (namely, the GDPR), to ensure proper respect for fundamental rights of individuals which may be affected by the PolicyCLOUD platform, including respect for individual privacy as an expression of a fundamental right and freedom. The principles and main rules within the GDPR, where applicable, have been developed in further detail within D3.3 and in D3.6, with an aim to identify practical requirements to be met by the platform.

D3.6 constitutes an update on D3.3, which was the first step of the legal, regulatory, societal, and ethical analysis. Starting from a more general perspective, D3.6 contributes to focus on the implementation of requirements to be applied in the PolicyCLOUD architecture, but also to its components and in different use cases scenarios. In detail:

- With regards to WP2 (“Requirements, Architecture & Innovation”), the main requirements envisaged are those related to data security and data retention.
- Concerning the requirements selected and discussed in the context of WP3 (“Cloud Infrastructure Utilization & Data Governance”) the attention is dedicated, firstly, to the cloud infrastructure itself and, secondly, to the data management, which is an issue that concerns both policy and end users’ constraints. Furthermore, in the context of this WP, the Interim Repository as well as the Incentives Management component have assumed importance at the point to become other elements of the legal, regulatory, societal, and ethical analysis.
- Specific requirements have also been addressed with regards to WP4 (“Reusable Models & Analytical Tools”), since this WP is mainly focused on analytics components (including registration of new analytics components), on data source registration and on data management, issues that concern both policymaking and end users’ constraints.
- Regarding WP5 (“Cross-sector Policy Lifecycle Management”), the key requirements are related to the PDT and the PME. Among these requirements, it is worth to be mentioned the implementation of a data protection information notice according with Art. 13 GDPR.
- In the context of WP6 (“Use Cases Adaptation, Integration & Experimentation”), the requirements constantly monitored and assessed in cooperation with the relevant Use Case Partners are those referred to the phase of the data source selection, that is more critical in scenarios C and D of use case 1 and in use case 2, since in the other scenarios and use cases no relevant personal data processing activities are being performed.

With reference to WP7 (“Communication, Exploitation, Standardisation, Roadmapping & Business Development”) the ethical, legal, regulatory, and societal assessment has been focused on the data marketplace. The identified requirements will need to be updated over the course of the Project’s development, to ensure that they remain aligned with the features and functionalities planned for the platform, as well as the specificities of the different use cases considered (e.g., data sources, scenarios, intended uses of the platform). As such, an effort will be made to simultaneously strive for the platform’s adherence to the identified requirements, and to adapt those requirements to ensure that they are

comprehensive, relevant and feasibly implementable within the platform. The Legal and Ethical Framework should thus be seen as evolving alongside the platform, with the end-goal of ensuring that the platform develops in a legally and ethically sound manner.

The different use cases may have different specific ethical implications. For this reason, in D3.3 and D3.6, a specific section for each use case has been provided and will be reviewed during the development of the project. What has been observed is that some ethical issues will arise in any of the situations described since the impacts relate to the type of technology used, meanwhile other ethical issues strictly rely on the features of the use case - e.g., in Use Case #1, the issues such as the transparency, the presence of bias are of particular importance, the “filter bubble” distortion may have bad impacts in the use cases related to the creation of public policies, since the fact to not consider properly the complexity and the 360-degree reality could be detrimental to some categories of subjects (in Use Case #3 or in Use Case #4) or objects (in Use Case #2).

In conclusion, beside the specific requirements that PolicyCLOUD should respect to be compliant with the regulatory framework, the ethical perspective, even though more intangible and less-countable must be respected and taken into account during – but foremost – after the project. For that reason, the Legal/Ethical Framework will need a continuous update. All the partners may contribute to produce a more coherent and satisfactory ethical framework, intervening to highlight critical points and situations arise during the testing phase or the deployment phase, to permit the addressing and resolutions of problems that may create discontinuity in the project.

Section	Description
ID	REQ- SY-LEF-01
Title	Legal/Ethical Framework
Description	The platform should be aligned with all applicable ethical, legal, regulatory and societal requirements
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Success Criteria	The platform and the use cases prove to be compliant to the compliance controls defined under T3.5
Expected delivery date	M36 (3rd version of prototypes)

TABLE 124 - SYSTEM REQUIREMENT REQ- SY-LEF-01

6 Software Requirements

This section provides an updated list of the software requirements for the PolicyCLOUD project. These requirements are related with specific software portions, which can be either a program, a software component, an existing product that will be used as part of the overall platform, or a set of combinations of all the above that implements a specific functionality and provides a set of capabilities via well-defined interfaces. They may include functional or non-functional requirements imposed by a specific software component that are related with:

- Interfaces exposed by the specific software component that describes the way of interaction with the other software portions
- Performance requirements upon this software portion
- Features that required to be implemented by other components
- Conditions or constraints that the software component should or must take into consideration

The following subsections contain all these software requirements per technological component that will provide an autonomous functionality. They will consist of a specific software building block in the overall PolicyCLOUD architecture.

6.1 Cloud Provisioning

Section	Description
ID	REQ- SO-CP-01
Title	Provisioning of cloud-based resources to set-up the PolicyCLOUD infrastructure
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	<p>The computing resources of the provisioned cloud infrastructure should be scalable to address the requirements of the selected use cases. To set-up the PolicyCLOUD infrastructure, the cloud provider should contribute with the following pledged resources:</p> <ul style="list-style-type: none"> • 68 vCPU cores, 308 GB of RAM • 2TB of block storage and 50-100MB of object storage • PaaS Orchestrator to access the resources • The first day of service delivery is Aug. 2020 • The last day of service delivery is Dec. 2022
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02
Source	N/A

Section	Description
Success Criteria	Partners will access the PolicyCLOUD infrastructure, operate a distributed K8s cluster as a service for the project, and deploy services and pilot use cases using the PaaS orchestrator made available by the cloud provider
Expected delivery date	A Service Level Agreement (SLA) with the customer (the consortium of PolicyCLOUD) and an Operational Level Agreement (OLA) with the cloud provider was agreed on Oct. 2020. This agreement is subject to review on an annual basis (at the end of the calendar year from the service delivery date) by which a customer performance validation will be conducted, and a report will be produced

TABLE 125 - SOFTWARE REQUIREMENT REQ - SO-CP-01

6.2 Cloud Register

Section	Description
ID	REQ- SO-CR-01
Title	Access the PolicyCLOUD IaaS and PaaS
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Access to the IaaS/PaaS will be available either via GUI or CLI
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02
Source	The registration process to access the cloud infrastructure provided by EGI is based on two steps: <ul style="list-style-type: none"> • Sign up the EGI User Community and join the vo.policycloud.eu organization • Validate and activate the registration to the EGI User Community
Success Criteria	Partners will access the PolicyCLOUD PaaS and IaaS with federated credentials
Expected delivery date	Access to the PolicyCLOUD PaaS and IaaS will be available from M06

TABLE 126 - SOFTWARE REQUIREMENT REQ - SO-CR-01

6.3 Cloud Gateways

Section	Description
ID	REQ- SO-CG-01
Title	Connection to APIs
Level of detail	Software
Type	FUN (Function)
Description	The PolicyCLOUD Gateway Component should facilitate the connection to appropriately specified APIs for the retrieval of the information, integrating the corresponding security measures and safeguarding information integrity
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful connection established to the defined data source
Expected delivery date	M12 (1st version of prototypes)

TABLE 127 - SOFTWARE REQUIREMENT REQ -SO-CG-01

Section	Description
ID	REQ- SO-CG-02
Title	File Parsing
Level of detail	Software
Type	FUN (Function)
Description	Parsing of files (e.g., excel or csv files) should be facilitated for the retrieval of the information, for integrating the corresponding security measures and for safeguarding information integrity
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful connection established to the defined data source
Expected delivery date	M12 (1st version of prototypes)

TABLE 128 - SOFTWARE REQUIREMENT REQ -SO-CG-02

Section	Description
ID	REQ- SO-CG-03
Title	Connection to (SQL or No-SQL) Databases
Level of detail	Software
Type	FUN (Function)
Description	The connection to an appropriately specified (SQL or No-SQL) database should be accomplished in order to achieve the retrieval of the information, the integration of the corresponding security measures, and the safeguarding of information integrity
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful connection established to the defined data source
Expected delivery date	M12 (1st version of prototypes)

TABLE 129 - SOFTWARE REQUIREMENT REQ -SO-CG-03

Section	Description
ID	REQ- SO-CG-04
Title	Configuration
Level of detail	Software
Type	FUN (Function)
Description	The Gateway Component should provide access to a configuration service, facilitating configuration of the connection parameters per connection type and source
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful configuration of the connection parameters
Expected delivery date	M12 (1st version of prototypes)

TABLE 130 - SOFTWARE REQUIREMENT REQ -SO-CG-04

Section	Description
---------	-------------

ID	REQ- SO-CG-05
Title	Pull Connection Type Support
Level of detail	Software
Type	FUN (Function)
Description	The Gateway Component should support pulling data from external data sources (e.g., through REST APIs) per predefined time intervals
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful retrieval of information from the defined data sources
Expected delivery date	M12 (1st version of prototypes)

TABLE 131 - SOFTWARE REQUIREMENT REQ -SO-CG-05

Section	Description
ID	REQ- SO-CG-06
Title	Push Connection Type Support
Level of detail	Software
Type	FUN (Function)
Description	The Gateway Component should support data from external data sources being pushed to the platform per predefined time intervals
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful collection and internal push of information and data from the defined data sources
Expected delivery date	M12 (1st version of prototypes)

TABLE 132 - SOFTWARE REQUIREMENT REQ -SO-CG-06

Section	Description
---------	-------------

ID	REQ- SO-CG-07
Title	Standardized Interface to other internal PolicyCLOUD components
Level of detail	Software
Type	FUN (Function)
Description	The Gateway Component should facilitate the standardised connection to other internal components of the PolicyCLOUD platform, such as the Data Cleaning Component, the Data Fusion Component, etc. The standardisation of the messages should follow a well-defined and structured format, such as XML or JSON
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Proper specification of message structure
Expected delivery date	M12 (1st version of prototypes)

TABLE 133 - SOFTWARE REQUIREMENT REQ -SO-CG-07

Section	Description
ID	REQ- SO-CG-08
Title	Global project's social media API keys and credentials
Level of detail	Software
Type	FUN (Function)
Description	API keys and credentials for accessing, integrating and connecting with widely used social media and platforms should be provided in order the Cloud Gateway to be able to fetch data from these specific external sources. A global and public project's account should be established in order to retrieve the needed keys and credentials
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful interconnection with external platforms through the utilization of their APIs
Expected delivery date	M12 (1st version of prototypes)

TABLE 134 - SOFTWARE REQUIREMENT REQ -SO-CG-08

Section	Description
---------	-------------

ID	REQ- SO-CG-09
Title	Define and establish an environment for setting up the component
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	A concrete and stable environment for hosting and providing the Cloud Gateway component should be established and provided. Needed resources of 1VM with 4GB RAM, 4 cores, and 40GB space should be facilitated
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	Proper installation and utilization of provided VM
Expected delivery date	M12 (1st version of prototypes)

TABLE 135 - SOFTWARE REQUIREMENT REQ -SO-CG-09

Section	Description
ID	REQ- SO-CG-10
Title	Create a docker image of the component
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Kubernetes cluster will be utilized for deploying the project's components. To this end, Cloud Gateway component will be provided as a docker image to be able to be deployed in Kubernetes
Additional Information	Cloud Gateway component could be executed and provided as a standalone software by installing it directly into a VM, just in case the Kubernetes platform is not available
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	A successful deployment using Kubernetes clustering
Expected delivery date	M24 (2nd version of prototypes)

TABLE 136 - SOFTWARE REQUIREMENT REQ -SO-CG-10

Section	Description
ID	REQ- SO-CG-11
Title	Connection with Kafka queues

Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	An interconnection with a Kafka Cluster should be facilitated. Streaming data fetching from streaming data source, like in the case of Twitter Streaming API, should be published and ingested in a Kafka cluster
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Streaming data should be consumed and ingested by the Kafka Cluster without interruptions and problems
Expected delivery date	M12 (1st version of prototypes)

TABLE 137 - SOFTWARE REQUIREMENT REQ -SO-CG-11

Section	Description
ID	REQ- SO-CG-12
Title	Integration with Keycloak
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	It has been identified the need to integrate Cloud Gateways component with the PolicyCLOUD's federated identity mechanism, Keycloak. To this end, users in order to obtain access in protected resources have to send along with every request, a token obtained from Keycloak that ensures the identity and role.
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	The Gateway authentication and security mechanism ensure that only authorized requests can get access to services and data
Expected delivery date	M24 (2nd version of prototypes)

TABLE 138 - SOFTWARE REQUIREMENT REQ -SO-CG-12

Section	Description
ID	REQ- SO-CG-13
Title	Integration with Interim Repository

Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Interim repository has been identified as project's temporary data storage that is being used to store information and datasets that have been uploaded from different use cases. Cloud Gateways should ensure the interconnection with this repository in order to fetch the wanted datasets
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Fetch all the wanted datasets from this persistent repository
Expected delivery date	M24 (2nd version of prototypes)

TABLE 139 - SOFTWARE REQUIREMENT REQ -SO-CG-13

Section	Description
ID	REQ- SO-CG-14
Title	Open API Documentation
Level of detail	Software
Type	FUN
Description	Cloud Gateway REST API should follow the Open API specifications. The Open API approach should be followed in order to be easier for the end user to discover the capabilities of the Cloud Gateway component and provide well-structured documentation for each of the component's services
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Include an API documentation page, by using Swagger UI, so that a graphical interface for interacting with the exposed APIs can be provided
Expected delivery date	REQ- SO-CG-14

TABLE 140 - SOFTWARE REQUIREMENT REQ -SO-CG-14

Section	Description
ID	REQ- SO-CG-15
Title	Integration with Apache Avro

Level of detail	Software
Type	FUN
Description	Avro serialization should be utilized for the use cases that handle large records as JSON serialization may have a negative impact for the performance and resource consumption of the microservice
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	UC#1, UC#3, UC#4
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Output data messages should be Avro-serialized before to be sent to Kafka topics.
Expected delivery date	M24 (2nd version of prototypes)

TABLE 141 - SOFTWARE REQUIREMENT REQ -SO-CG-15

Section	Description
ID	REQ- SO-CG-16
Title	Integration with Traefik
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Traefik instead of requiring manual route configuration for each service component, bundles to the registry service or orchestrator API and generates all routes automatically so this services to be available for public and ready to use. Continuously updating its configurations can be really helpful in a microservices environment that changes are common issues and component restarts are required
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Multiple nodes of the same services must be deployed when needed offering a load balancing solution to the Cloud Gateway component
Expected delivery date	M24 (2nd version of prototypes)

TABLE 142 - SOFTWARE REQUIREMENT REQ -SO-CG-16

6.4 Incentives Management

Section	Description
---------	-------------

ID	REQ- SO-IM-01
Title	Define an interface for the component to set the incentives
Level of detail	Software
Type	FUNC (function)
Description	An interface will be developed to set different fields that instantiate an incentive defined by the policy makers and should be executed by the PDT
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-03
Source	N/A
Success Criteria	An incentive can be defined in PolicyCLOUD
Expected delivery date	M24 (2nd version of prototypes)

TABLE 143 - SOFTWARE REQUIREMENT REQ - SO-IM-01

Section	Description
ID	REQ- SO-IM-02
Title	Being able to manage the incentives defined in PolicyCLOUD
Level of detail	Software
Type	DATA (data)
Description	The incentives defined should be stored in the PolicyCLOUD datastore in a concrete schema. Moreover, those incentives should be managed and consulted by the PDT
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-03
Source	N/A
Success Criteria	Incentives can be managed in PolicyCLOUD
Expected delivery date	M24 (2nd version of prototypes)

TABLE 144 - SOFTWARE REQUIREMENT REQ - SO-IM-02

6.5 Data Cleaning

Section	Description
---------	-------------

ID	REQ- SO-DC-01
Title	Standardised Interface to other internal PolicyCLOUD components
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should facilitate the standardised connection to other internal components of the PolicyCLOUD platform, such as the Cloud Gateways. The standardisation of the messages should follow the JSON well-defined and structured format
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python library of JSON
Success Criteria	Proper specification of message structure
Expected delivery date	M12 (1st version of prototypes)

TABLE 145 - SOFTWARE REQUIREMENT REQ- SO-DC-01

Section	Description
ID	REQ- SO-DC-02
Title	Error identification
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should facilitate the identification of errors associated with conformance to specific set constraints
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 146 - SOFTWARE REQUIREMENT REQ- SO-DC-02

Section	Description
ID	REQ- SO-DC-03
Title	Conformance to specific data types

Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to specific data types (e.g., integer, string, etc.)
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 147 - SOFTWARE REQUIREMENT REQ- SO-DC-03

Section	Description
ID	REQ- SO-DC-04
Title	Conformance to range constraints
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to specific range constraints (e.g., min and max values)
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 148 - SOFTWARE REQUIREMENT REQ- SO-DC-04

Section	Description
ID	REQ- SO-DC-05
Title	Conformance to predefined values
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to specific predefined values (e.g., values selected from a drop-down list)
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 149 - SOFTWARE REQUIREMENT REQ- SO-DC-05

Section	Description
ID	REQ- SO-DC-06
Title	Conformance to regular expression patterns
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to regular expression patterns (data that has a certain pattern in the way it is displayed, such as phone numbers (e.g., for text formatting “123-45-6789” or “123456780” or “123 45 6789”))
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 150 - SOFTWARE REQUIREMENT REQ- SO-DC-06

Section	Description
ID	REQ- SO-DC-07
Title	Conformance to value separation
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to separation of values (e.g., complete address in free form field without any indication where street ends, and city begins)
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M24 (2nd version of prototypes)

TABLE 151 - SOFTWARE REQUIREMENT REQ- SO-DC-07

Section	Description
ID	REQ- SO-DC-08
Title	Conformance to cross-field validity
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to cross-field validity (e.g., the sum of the parts of data must equal to a whole)
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M24 (2nd version of prototypes)

TABLE 152 - SOFTWARE REQUIREMENT REQ- SO-DC-08

Section	Description
ID	REQ- SO-DC-09
Title	Conformance to correct value representation
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to correct representation of the values
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 153 - SOFTWARE REQUIREMENT REQ- SO-DC-09

Section	Description
ID	REQ- SO-DC-10
Title	Conformance to uniqueness
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard the conformance of the data to their uniqueness (i.e., data that cannot be repeated since they require unique values (e.g., social security numbers))
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 154 - SOFTWARE REQUIREMENT REQ- SO-DC-10

Section	Description
ID	REQ- SO-DC-11
Title	Conformance to mandatory field
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that all the mandatory fields of the data are filled in
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 155 - SOFTWARE REQUIREMENT REQ- SO-DC-11

Section	Description
ID	REQ- SO-DC-12
Title	Conformance to specific value length
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that all the filled in values that have specific length constraints, are correctly placed
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 156 - SOFTWARE REQUIREMENT REQ- SO-DC-12

Section	Description
ID	REQ- SO-DC-13
Title	Conformance to specific coding standard
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that all the appropriate attributes of the data respect and apply their predefined coding standard
Additional Information	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M24 (2nd version of prototypes)

TABLE 157 - SOFTWARE REQUIREMENT REQ- SO-DC-13

Section	Description
ID	REQ- SO-DC-14
Title	Conformance to value uniformity
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that the appropriate attributes of the data respect and follow their predefined value representation
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 158 - SOFTWARE REQUIREMENT REQ- SO-DC-14

Section	Description
ID	REQ- SO-DC-15
Title	Identification of duplications
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should facilitate the identification of duplications across the data
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 159 - SOFTWARE REQUIREMENT REQ- SO-DC-15

Section	Description
ID	REQ- SO-DC-16
Title	Automatic field completion
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that the data is fully complete
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Automatic completion of on-purpose excluded values
Expected delivery date	M24 (2nd version of prototypes)

TABLE 160 - SOFTWARE REQUIREMENT REQ- SO-DC-16

Section	Description
ID	REQ- SO-DC-17
Title	Automatic error correction
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that inconsistencies and errors that exist on the data are identified and properly corrected
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	Automatic correction of on-purpose included erroneous values
Expected delivery date	M24 (2nd version of prototypes)

TABLE 161 - SOFTWARE REQUIREMENT REQ- SO-DC-17

Section	Description
ID	REQ- SO-DC-18
Title	Data verification
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should safeguard that the data is accurate, especially referring to erroneous inliers (i.e., data points generated by error but falling within the expected range (erroneous inliers often escape detection))
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python library of Cerberus
Success Criteria	Identification of on-purpose included errors
Expected delivery date	M12 (1st version of prototypes)

TABLE 162 - SOFTWARE REQUIREMENT REQ- SO-DC-18

Section	Description
ID	REQ- SO-DC-20
Title	Integration with Apache Kafka queue
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should be able to send and retrieve data from different Apache Kafka queues, in order to exchange all the needed data with the interconnected components (i.e., Cloud Gateways and Data Interoperability)
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python library of Kafka
Success Criteria	Successful establishment of a connection with a Kafka queue
Expected delivery date	M24 (2nd version of prototypes)

TABLE 163 - SOFTWARE REQUIREMENT REQ- SO-DC-20

Section	Description
ID	REQ- SO-DC-21
Title	Registration to OpenWhisk
Level of detail	Software
Type	FUNC (function)
Description	The Data Cleaning should be registered to OpenWhisk in order to be triggered and be utilized as an analytic component
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with OpenWhisk
Expected delivery date	M24 (2nd version of prototypes)

TABLE 164 - SOFTWARE REQUIREMENT REQ- SO-DC-21

6.6 Data Fusion Linking

Section	Description
ID	REQ- SO-DFL-01
Title	Kubernetes, OpenWhisk and Spark clusters
Level of detail	System
Type	ENV (Operational/Environment Requirements)
Description	A Kubernetes cluster and a Spark cluster installed on it. The actual hardware requirements for a functional PolicyCLOUD system depends on the amount of expected data to be ingested and analysed by the system. For the project development, the requirement is of 4 VMs each with 16 cores and 128GB memory.
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ALL
Source	Kubernetes, OpenWhisk, Apache Spark
Success Criteria	Demonstration of use cases with large amount of data, with reasonable performance
Expected delivery date	Initial version on M12, enhancements on M24, M36

TABLE 165 - SOFTWARE REQUIREMENT REQ - SO-DFL-01

Section	Description
ID	REQ- SO-DFL-02
Title	Data streaming framework with initial analytic during data ingest
Level of detail	Software
Type	FUNC (function)
Description	A scalable data streaming middleware framework (based on serverless platform of openWhisk and Kafka) for integrating analytic functions to process the ingested data
Additional Information	The analytic functions will be applied according to registered specification per data source
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ALL
Source	Apache Spark Streaming
Success Criteria	Demonstration of processing and initial analytic on a registered data source providing value to the policy use case
Expected delivery date	Initial version on M12, enhancements on M24, M36

TABLE 166 - SOFTWARE REQUIREMENT REQ - SO-DFL-02

Section	Description
ID	REQ- SO-DFL-03
Title	Data source & tool registration for streaming analytic
Level of detail	Software
Type	FUNC (function)
Description	A capability to register analytic function and register a data source (with schema / metadata) for streaming analytic by a registered analytic function(s) that support the schema / metadata, and applying the registered analytics during streaming of that data source
Additional Information	The registration will include parameters for the analytic function(s)
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-01, ROL-02
Source	Registration analytic tasks for activation in Apache Spark Streaming
Success Criteria	Demonstration of registration of new analytic functions, data source the use it, and policy validation scenario using the function and data source
Expected delivery date	Initial version on M24, enhancements on M36

TABLE 167 - SOFTWARE REQUIREMENT REQ - SO-DFL-03

Section	Description
ID	REQ- SO-DFL-04
Title	Data source & tool registration for regular analytics on data at rest
Level of detail	Software
Type	FUNC (function)
Description	A capability to register analytic function and register a data source (with schema / metadata) that can be a subject to a regular analytic on data at rest (that was already ingested to the system) by a registered analytic function(s) that support the schema / metadata
Additional Information	The registration will include parameters for the analytic function(s)
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-01, ROL-02
Source	N/A
Success Criteria	Demonstration of registration of new analytic functions, data source they use it, and policy validation scenario using the function and data source
Expected delivery date	Initial version on M24, enhancements on M36

TABLE 168 - SOFTWARE REQUIREMENT REQ - SO-DFL-04

Section	Description
ID	REQ- SO-DFL-05
Title	Seamless Analytics on Hybrid Data at Rest
Level of detail	Software
Type	FUNC (function)
Description	Capability of applying analytics seamlessly on data on multiple stores, and mechanism to move older data to long term store
Additional Information	Specifically, newer (hot) data will be ingested into database while older data will be moved periodically to object storage
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02
Source	N/A
Success Criteria	Demonstration of seamless analytic with the data movement process
Expected delivery date	M24 (2nd version of prototypes)

TABLE 169 - SOFTWARE REQUIREMENT REQ - SO-DFL-05

Section	Description
ID	REQ- SO-DFL-06
Title	Data privacy and ownership constraints for multi-tenant analytics
Level of detail	Software
Type	FUNC (function)
Description	The Data Acquisition and Analytics Layer should maintain access control mechanism to respect data privacy stings of the data owner
Additional Information	Applying analytics to data should be restricted according to the data privacy settings of the data owner
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	Validation test of data access restrictions
Expected delivery date	M24 (2nd version of prototypes)

TABLE 170 - SOFTWARE REQUIREMENT REQ - SO-DFL-06

6.7 Data Interoperability

Section	Description
ID	REQ- SO-DI-01
Title	Cleaned Data
Level of detail	Software
Type	DATA (data)
Description	The Data Interoperability component extracts semantic knowledge and good quality information from the cleaned data that will be the input to its system
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	Development of good quality annotated and interoperable data from the provided cleaned data
Expected delivery date	M12 (1st version of prototypes)

TABLE 171 - SOFTWARE REQUIREMENT REQ - SO-DI-01

Section	Description
ID	REQ- SO-DI-02
Title	Triplestore Database
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Triplestore is needed in order to save correlated, annotated and interoperable data in JSON-LD format and as linked ontologies. Hence, it will be feasible the storage of semantic facts and the support of the corresponding data schema models
Additional Information	Apache JENA is the preferred Triplestore framework to be used
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	Apache Jena Framework
Success Criteria	Successful saving of interoperable data in JSON-LD formats and as linked ontologies
Expected delivery date	M12 (1st version of prototypes)

TABLE 172 - SOFTWARE REQUIREMENT REQ - SO-DI-02

Section	Description
ID	REQ- SO-DI-03
Title	Data Modelling & Ontology Mapping
Level of detail	Software
Type	FUNC (function)
Description	Define the appropriate techniques and tools to map concepts, classes, and semantics defined in different ontologies and datasets and to achieve transformation compatibility through extracted metadata
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	Successful annotation, transformation and mapping of data and corresponding ontologies in terms of semantic and syntactic interoperability of data
Expected delivery date	M12 (1st version of prototypes)

TABLE 173 - SOFTWARE REQUIREMENT REQ - SO-DI-03

Section	Description
ID	REQ- SO-DI-04
Title	Data Schemas & Data Models
Level of detail	Software
Type	DATA (data)
Description	Define the exact data schemas and models that will be used from the analytical components and will derive and produced by the Data Interoperability Component. Incoming and cleaned data will be modelled and transformed according to the defined schemas and models
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-03, ROL-04
Source	N/A
Success Criteria	Development of models and schemas corresponding to reference and analytical problems and tasks.
Expected delivery date	M12 (1st version of prototypes)

TABLE 174 - SOFTWARE REQUIREMENT REQ - SO-DI-04

Section	Description
ID	REQ- SO-DI-05
Title	Entities identification and recognition
Level of detail	Software
Type	DATA (data)
Description	Define the entities that will be used for interconnection and interlinking with widely used knowledge bases. Classifying named entities found in cleaned data into pre-defined categories, such as persons, places, organizations, dates, etc
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Spacy, Scikit-learn
Success Criteria	Successful identification of entities from provided cleaned data
Expected delivery date	M12 (1st version of prototypes)

TABLE 175 - SOFTWARE REQUIREMENT REQ - SO-DI-05

Section	Description
ID	REQ- SO-DI-06
Title	Entities Interlinking
Level of detail	Software
Type	DATA (data)
Description	Annotate cleaned data with URIs pointing into corresponding Wikidata and DBPedia entities & proper Ontologies
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Spacy, PWikidata, DBPedia Spotlight
Success Criteria	Successful and proper interlinking of the recognised entities from provided cleaned data
Expected delivery date	M12 (1st version of prototypes)

TABLE 176 - SOFTWARE REQUIREMENT REQ - SO-DI-06

Section	Description
ID	REQ- SO-DI-07
Title	Automatic topic identification
Level of detail	Software
Type	FUNC (function)
Description	The Data Interoperability mechanism should safeguard that datasets' topic of interest are proper identified. Proper topic analysis should be facilitated in order to organize and understand large collections of text data
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Scikit-learn, Spacy and Gensim
Success Criteria	Automatic topic identification of provided datasets
Expected delivery date	M24 (2nd version of prototypes)

TABLE 177 - SOFTWARE REQUIREMENT REQ - SO-DI-07

Section	Description
ID	REQ- SO-DI-08
Title	Integration with Apache Kafka queue
Level of detail	Software
Type	FUNC (function)
Description	It has been identified that data should be ingested and flow between different project's components through the utilization of Kafka queues. Cleaned data should be consumed by the Interoperability mechanism from a Kafka consumer, and then transformed and finally ingested into a Kafka producer and another Kafka queue, preserving data consistency
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful establishment of a connection with a Kafka queue
Expected delivery date	M24 (2nd version of prototypes)

TABLE 178 - SOFTWARE REQUIREMENT REQ - SO-DI-08

Section	Description
ID	REQ- SO-DI-09
Title	Registration to OpenWhisk framework
Level of detail	Software
Type	FUNC (function)
Description	It has been identified the usage of OpenWhisk framework for triggering and utilizing analytic components through APIs. To this end, the interoperability component should be integrate and register to OpenWhisk framework, in order to be provided as a service through a corresponding API
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the OpenWhisk framework
Expected delivery date	M24 (2nd version of prototypes)

TABLE 179 - SOFTWARE REQUIREMENT REQ - SO-DI-09

6.8 Data Store

Section	Description
ID	REQ- SO-DS-01
Title	Minimum hardware requirements
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	The datastore requires minimum 4G of memory to operate
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The datastore should be running with no problems
Expected delivery date	M12 (1st version of prototypes)

TABLE 180 - SOFTWARE REQUIREMENT REQ - SO-DS-01

Section	Description
ID	REQ- SO-DS-02
Title	Fragment a dataset and move the data fragments across different nodes
Level of detail	Software
Type	DATA (data)
Description	The adaptable distributed storage must split a dataset into different regions, and move these regions to diverse data nodes to adapt in case of increased load (user workload/data load), achieving efficient consumption based on provided resources
Additional Information	When a data fragment movement (move, split, join) occurs, the storage must not suffer from down-time remaining operational with minimum performance overhead
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Data can be moved in different nodes
Expected delivery date	M24 (2nd version of prototypes)

TABLE 181 - SOFTWARE REQUIREMENT REQ - SO-DS-02

Section	Description
ID	REQ- SO-DS-03
Title	Provide standard connectivity mechanisms
Level of detail	Software
Type	DATA (data)
Description	The datastore must implement standard connectivity mechanisms to provide access and allow for the query execution
Additional Information	Provide JDBC ⁷ implementation, and as additional standard implementation OData ⁸
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	All components can retrieve and store data to the data store
Expected delivery date	M24 (2nd version of prototypes)

⁷ <https://jcp.org/en/jsr/detail?id=221>

⁸ <https://www.odata.org/>

TABLE 182 - SOFTWARE REQUIREMENT REQ - SO-DS-03

Section	Description
ID	REQ- SO-DS-04
Title	Requirement for a Kubernetes cluster to enable the deployment
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	The infrastructure of PolicyCLOUD should use Kubernetes for deploying the various application/platform components, the adaptable distributed engine must be able to deploy and configure additional data nodes via this technology
Additional Information	N/A
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	A successful deployment using Kubernetes clustering
Expected delivery date	M12 (1st version of prototypes)

TABLE 172 - SOFTWARE REQUIREMENT REQ - SO-DS-04

Section	Description
ID	REQ- SO-DS-05
Title	Support data ingestion in high rates from a stream
Level of detail	Software
Type	DATA (data)
Description	It has been identified the need to ingest data coming from a data stream to the data store, preserving data consistency in terms of database transactions
Additional Information	Data ingestion at high rates cannot be supported by traditional operational databases that apply microbatches. Also, operational datastores cannot allow analytics on top of operational data, as the analytics conflict with the transactional processing. The PolicyCLOUD datastore must allow data ingestions in high rates and in the meantime, providing efficient execution of analytics, so that the tools can use the real data, without the need to export the data into an additional data warehouse
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Similar response time when performing analytics with or without data ingestion
Expected delivery date	M12 (1st version of prototypes)

TABLE 183 - SOFTWARE REQUIREMENT REQ - SO-DS-05

Section	Description
ID	REQ- SO-DS-06
Title	Integration with Apache Kafka ⁹ queue
Level of detail	Software
Type	DATA (data)
Description	It has been identified the need to ingest data coming from a data stream to the data store, preserving data consistency in terms of database transactions
Additional Information	The use of Kafka queues will interconnect the various components that formulate a complete data pipeline. Thus, at the end of the pipeline data will be pushed to a queue, where the datastore should retrieve them and persistently store. There is a need for a data connectivity mechanism with the queue
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Establish a connection with Kafka queue and allow data insertions. Data incoming to the queue must be then stored in the data store
Expected delivery date	M24 (2nd version of prototypes)

TABLE 184 - SOFTWARE REQUIREMENT REQ - SO-DS-06

Section	Description
ID	REQ- SO-DS-07
Title	Transparent ingest data using a well-defined data schema
Level of detail	Software
Type	DATA (data)
Description	Given that the data provider defines its data schema in well-known format (i.e., Apache Avro Registry) the data must be imported transparently, to the datastore, as soon as it appears in a Kafka topic
Additional Information	N/A
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Establish a connection with Kafka queue and allow data insertions. Data incoming to the queue must be then stored in the data store

⁹ <http://kafka.apache.org/>

Expected delivery date	M24 (2nd version of prototypes)
-------------------------------	---------------------------------

TABLE 185 - SOFTWARE REQUIREMENT REQ - SO-DS-07

6.9 Opinion Mining

Section	Description
ID	REQ- SO-OM-01
Title	Opinion Mining
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe events and social attitude regarding specifics topics (i.e., a policy, a demonstration, a group of people, a wine) extracted from datasets and social networks
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	Capturean tool
Success Criteria	The opinion mining component should be running with no problems
Expected delivery date	M24 (2nd version of prototypes)

TABLE 186 - SOFTWARE REQUIREMENT REQ - SO-OM-01

Section	Description
ID	REQ- SO-OM-03
Title	Social media graph analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to identify those users who are talking more about a topic (i.e., a policy, a demonstration, a group of people, a wine, et.c)
Additional Information	N/A
Priority	DES (desirable requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The opinion mining component should be running with no problems

Expected delivery date	M36 (3 rd version of prototypes)
-------------------------------	---

TABLE 187 - SOFTWARE REQUIREMENT REQ - SO-OM-03

Section	Description
ID	REQ- SO-OM-04
Title	Twitter User Monitoring
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to identify and monitor most popular users (at least on Twitter) who comment about specifics hashtags or topics
Additional Information	Their provision will depend on the use case's needs
Priority	OPT (optional requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The opinion mining component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 188 - SOFTWARE REQUIREMENT REQ - SO-OM-04

Section	Description
ID	REQ- SO-OM-07
Title	Support the integration with LXS storage
Level of detail	Software
Type	FUNC (function)
Description	The components developed to fulfil the requirements included in this section should be able to feed At-Rest data from LXS data storage and dump results back
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the LXS Data Storage
Expected delivery date	M24 (2nd version of prototypes), M36 (3 rd version of prototypes)

TABLE 189 - SOFTWARE REQUIREMENT REQ - SO-OM-07

Section	Description
ID	REQ- SO-DI-08
Title	Registration to OpenWhisk framework
Level of detail	Software
Type	FUNC (function)
Description	The Opinion Mining component should be integrate and register to OpenWhisk framework, in order to be provided as a service through a corresponding API.
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the OpenWhisk framework
Expected delivery date	M24 (2nd version of prototypes) and M36 (3 rd version of prototypes)

TABLE 190 - SOFTWARE REQUIREMENT REQ - SO-OM-08

Section	Description
ID	REQ- SO-OM-09
Title	Trend analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker can observe solid or trend results based on historic information
Additional Information	e.g., in UC#1 - Keyword detection, new entities and terms recognition, or discover hidden patterns and new tendencies. It a specific type of predictive analysis
Priority	OPT (optional requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	The situational knowledge acquisition component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 191 - SOFTWARE REQUIREMENT REQ - SO-OM-09

6.10 Sentiment Analysis

Section	Description
ID	REQ- SO-SA-01
Title	Social Media Sentiment Analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe the sentiment about what the citizens say in social media channels regarding certain topics
Additional Information	A document-level approach will be followed for this requirement
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The sentiment component should be running with no problems
Expected delivery date	M12 (1 st version of prototypes)

TABLE 192 - SOFTWARE REQUIREMENT REQ - SO-SA-01

Section	Description
ID	REQ- SO-SA-02
Title	RSS Feed Sentiment Analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe the sentiment in RSS feeds channels regarding certain topics
Additional Information	N/A
Priority	DES (desirable requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The sentiment component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 193 - SOFTWARE REQUIREMENT REQ- SO-SA-02

Section	Description
---------	-------------

ID	REQ- SO-SA-03
Title	Enhanced Social Media Sentiment Analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe the sentiment towards what the citizens say in social media regarding certain entities previously identified in a text
Additional Information	This entity-level sentiment analysis will be provisioned in collaboration with REQ- SO-DI-05
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The sentiment component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 194 - SOFTWARE REQUIREMENT REQ- SO-SA-03

Section	Description
ID	REQ- SO-SA-04
Title	Support the integration with LXS storage
Level of detail	Software
Type	FUNC (function)
Description	The components developed to fulfil the requirements included in this section should be able to feed At-Rest data from LXS data storage and dump results back
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the LXS Data Storage
Expected delivery date	M24 (2nd version of prototypes), M36 (3 rd version of prototypes)

TABLE 195 - SOFTWARE REQUIREMENT REQ- SO-SA-04

Section	Description
ID	REQ- SO-SA-05
Title	Registration to OpenWhisk framework
Level of detail	Software
Type	FUNC (function)
Description	The sentiment analysis component should be integrate and register to OpenWhisk framework, in order to be provided as a service through a corresponding API
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the OpenWhisk framework
Expected delivery date	M24 (2nd version of prototypes) and M36 (3 rd version of prototypes)

TABLE 196 - SOFTWARE REQUIREMENT REQ- SO-SA-05

6.11 Behavioural Analysis

Section	Description
ID	REQ- SO-BA-01
Title	Policy modelling language
Level of detail	Software
Type	FUNC (function)
Description	A special-purpose modelling language needs to be developed that will allow policy makers to describe the characteristics of the population on which the policy will be applied and the specific policy mechanisms
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-03, ROL-04
Source	Code written in Elixir & JavaScript residing in: http://epinoetic.org/Politika/politika_demo.tar.gz
Success Criteria	Development of models corresponding to reference problems in network science
Expected delivery date	M12 (1st version of prototypes)

TABLE 197 - SOFTWARE REQUIREMENT REQ - SO-BA-01

Section	Description
ID	REQ- SO-BA-02
Title	Behaviour simulator
Level of detail	Software
Type	FUNC (function)
Description	Behavioural simulator that accepts as input and runs models developed using REQ-SO-BA-01
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-03, ROL-04
Source	Code written in Eixir & Behaviour residing in: http://epinoetic.org/Politika/politika_demo.tar.gz
Success Criteria	Execution of models created in REQ-SO-BA-02
Expected delivery date	M12 (1st version of prototypes)

TABLE 198 - SOFTWARE REQUIREMENT REQ - SO-BA-02

Section	Description
ID	REQ- SO-BA-03
Title	User Interface for the Behavioural Analysis component
Level of detail	Software
Type	FUNC (function), L&F (look & feel)
Description	Web-based interface that will allow I/O of population data and policy models along with control of the behavioural analysis component
Additional Information	Prototype resides in http://epinoetic.org/Politika/politika_demo.tar.gz
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-03, ROL-04
Source	Code written in Eixir & Behaviour residing in: http://epinoetic.org/Politika/politika_demo.tar.gz
Success Criteria	Web-based interaction with the REQ-SO-BA-02
Expected delivery date	M12 (1st version of prototypes)

TABLE 199 - SOFTWARE REQUIREMENT REQ - SO-BA-03

Section	Description
ID	REQ- SO-BA-04
Title	Fault-tolerant and safe operation of the behavioural analysis component
Level of detail	Software
Type	FUNC (function)
Description	Development of prevention, monitoring, and recovery methods for fault-tolerant and safe operation of the behavioural analysis component
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-02, ROL-03, ROL-04
Source	Code written in Elixir
Success Criteria	Stress testing the operation of the behavioural analysis component under extreme load and malicious/unsafe usage scenarios
Expected delivery date	M24 (2nd version of prototypes)

TABLE 200 - SOFTWARE REQUIREMENT REQ - SO-BA-04

Section	Description
ID	REQ- SO-BA-05
Title	Integration as an external component in PolicyCLOUD using the analytics tool registration & communication protocol procedures.
Level of detail	Software
Type	ENV (Operational/Environment Requirements)
Description	Development of an interface that allows communication of Behavioral Analysis as an external component with the rest of PolicyCLOUD
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	UC#1, UC#2, UC#3, UC#4
Role	ROL-01, ROL-02, ROL-04
Source	Code written in Elixir
Success Criteria	Storage of the simulation results of behavioural analysis into the PolicyCLOUD data store
Expected delivery date	M24 (2nd version of prototypes)

TABLE 201 - SOFTWARE REQUIREMENT REQ - SO-BA-05

6.12 Situational Knowledge Analysis

Section	Description
ID	REQ- SO-SKA-01
Title	Social Media Data Categorization
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe text-based information (coming from social media) classified into defined categories for report generation
Additional Information	The defined hierarchical classification (categories) will be provided by the use cases
Priority	OPT (optional requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The situational knowledge acquisition component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 202 - SOFTWARE REQUIREMENT REQ - SO-SKA-01

Section	Description
ID	REQ- SO-SKA-02
Title	RSS Feed data categorization
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe text-based information (coming from RSS Feed) classified into defined categories for report generation
Additional Information	The defined hierarchical classification (categories) will be provided by the use cases
Priority	OPT (optional requirement)
Reference Use Case	UC#1, UC#2
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	The situational knowledge acquisition component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 203 - SOFTWARE REQUIREMENT REQ - SO-SKA-02

Section	Description
ID	REQ- SO-SKA-03
Title	Supervised predictive analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to observe predictions based on historic information
Additional Information	e.g., the UC#2 - Prediction the quality of the next wine crop
Priority	MAN (mandatory requirement)
Reference Use Case	UC#2, UC#4
Role	ROL-01, ROL-02, ROL-04
Source	N/A
Success Criteria	The situational knowledge acquisition component should be running with no problems
Expected delivery date	M36 (3 rd version of prototypes)

TABLE 204 - SOFTWARE REQUIREMENT REQ - SO-SKA-03

Section	Description
ID	REQ- SO-SKA-04
Title	Descriptive analysis
Level of detail	Software
Type	FUNC (function)
Description	The policy maker shall be able to appropriately describe the characteristics of the dataset(s) of his/her interests and thus summarize them
Additional Information	This analysis can be found in the deliverables and/or prototypes under the name of Exploratory Data Analysis (EDA)
Priority	MAN (mandatory requirement)
Reference Use Case	UC#2, UC#4, UC#3
Role	ROL-01, ROL-02, ROL-04
Source	Python libraries of Pandas, NumPy, Scikit-learn, Keras
Success Criteria	The situational knowledge acquisition component should be running with no problems
Expected delivery date	M12 (1 st version of prototype) and M24 (2 nd version of prototypes)

TABLE 205 - SOFTWARE REQUIREMENT REQ - SO-SKA-04

Section	Description
ID	REQ- SO-SKA-05
Title	Being able to support the integration with LXS storage
Level of detail	Software
Type	FUNC (function)
Description	The components developed to fulfil the requirements included in this section should be able to feed At-Rest data from LXS data storage and dump results back
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the LXS Data Storage
Expected delivery date	M12 (1 st version of prototypes), M24 (2nd version of prototypes), M36 (3 rd version of prototypes)

TABLE 206 - SOFTWARE REQUIREMENT REQ - SO-SKA-05

Section	Description
ID	REQ- SO-SKA-06
Title	Registration to OpenWhisk framework
Level of detail	Software
Type	FUNC (function)
Description	The situational knowledge analysis component should be integrate and register to OpenWhisk framework, in order to be provided as a service through a corresponding API.
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-02, ROL-04
Source	N/A
Success Criteria	Successful integration with the OpenWhisk framework
Expected delivery date	M24 (2nd version of prototypes) and M36 (3 rd version of prototypes)

TABLE 207 - SOFTWARE REQUIREMENT REQ - SO-SKA-06

6.13 Optimization and Reusability

Section	Description
ID	REQ- SO-OR-01
Title	Use an operational and an analytical database to optimize the query execution
Level of detail	Software
Type	DATA (data)
Description	Data should be ingested to the operational datastore. When they become obsolete and thus, should be considered historical, they would need to be moved to the analytical datastore that can execute queries on BigData more efficiently
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Data can be moved from one database to the other and the performance should be improved
Expected delivery date	M24 (2nd version of prototypes)

TABLE 208 - SOFTWARE REQUIREMENT REQ - SO-OR-01

Section	Description
ID	REQ- SO-OR-02
Title	Provide access to data stores via a single and common interface
Level of detail	Software
Type	FUNC (function)
Description	PolicyCLOUD includes two different data stores - the LeanXcale relational data store and IBM object store. The dataset can be fragmented and distributed over the two data stores (historical data being moved to object store). However, the application should be kept unaware of these internal data transfers. The application needs a common interface to submit queries, without having to specify where the data is stored
Additional Information	A federation mechanism is required that will encapsulate the process of data retrieval from the two data stores. The LeanXcale access point will act as the federator between the relational and the Object Storage. The LeanXcale database already provides a common JDBC interface for data connectivity. The federator will receive the query and execute it in both data stores. For the object store, the access would be via Spark SQL, with the assistance of Apache Hive for storing the metadata of the schema catalogue, which can also be transparently accessible via a JDBC interface. The federator will take into consideration the operations that can be supported in order to push down the operations accordingly. Regarding the relational store, all operations will be pushed

Section	Description
	down to the store. At the very end, the federator will merge the results and return back the result set. It shouldn't count data that appears in both data stores twice
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	A query in the common interface can access data that are stored in both stores
Expected delivery date	M24 (2nd version of prototypes)

TABLE 209 - SOFTWARE REQUIREMENT REQ - SO-OR-02

Section	Description
ID	REQ- SO-OR-03
Title	Move historical data from the relational data store to the object store
Level of detail	Software
Type	DATA (data)
Description	Data ingested by the use cases will be stored into the relational datastore, as they are operational, in order to ensure data consistency in terms of ACID properties. After a configurable period, called the freshness window (which depends on the data set), the data becomes outdated and is no longer used by operational workloads. However, this historical data is still valuable and can be exploited by Big Data analytics algorithms. This data should be moved from the LeanXcale data base to the IBM object store
Additional Information	A mechanism should be implemented that monitors the freshness window and decides whether or not a data movement should take place. The mechanism must allow the data pulling of the data slice from the operational datastore and the persistently storage on the object store. During the data movement, the mechanism should allow the continuous execution of data retrieval from the data federator, so that no down time should be observed, while ensuring the data consistency
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Data can be moved automatically from one store to the other
Expected delivery date	M24 (2nd version of prototypes)

TABLE 210 - SOFTWARE REQUIREMENT REQ - SO-OR-03

Section	Description
ID	REQ- SO-OR-04

Section	Description
Title	Inform the LeanXcale data store when data are imported to the object store
Level of detail	Software
Type	FUNC (function)
Description	When data are pulled from the operational datastore, the LeanXcale data base can drop them. However, due to the asynchronous design, the LeanXcale data base cannot know when the data has been made available to the object store. As a result, the object store must inform the LeanXcale data base regarding the successful insertion of the data, so that the LeanXcale data base can safely drop these data
Additional Information	One possible solution to deal with this requirement will be the introduction of marking the data to be transferred to the object store by additional timestamps. Data that is being flushed and exported to the object store can be marked that way, so that later, the object store can inform the LeanXcale data base that this bunch of data has been successfully imported. By doing so, the federator component can push down operations accordingly, and only request specific data from the underlying data stores. Data that are known to the LeanXcale database that has been previously uploaded to the object store, will not be retrieved by the federator and can be safely discarded by the vacuum process of the LeanXcale data base
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Data moved to object store are now dropped from LeanXcale. A query to LeanXcale will not return any results regarding those data
Expected delivery date	M24 (2nd version of prototypes)

TABLE 211 - SOFTWARE REQUIREMENT REQ- SO-OR-04

Section	Description
ID	REQ- SO-OR-05
Title	Optimize query execution
Level of detail	Software
Type	DATA (Data)
Description	The federator receives a query and executes it into the different stores. The federator will be based on the LeanXcale query engine. The latter provides a query optimizer, which allows it to examine the different execution plans that can be produced in order to execute a query. However, it has been implemented to evaluate plans to be executed locally. It should be extended in order to take into consideration the operations that can be pushed down to the object store, and whether or not it is worth for an operator to be pushed down, according to the response time of the execution from Spark SQL, the amount of data that will be retrieved to the federator etc.
Additional Information	As every operation that can be supported by the object store will be pushed down to be executed locally, in order to avoid transferring a big amount of data through the

Section	Description
	network and process them in the query engine level, the implementation of this requirement corresponds to the following two aspects - the choose of the optimal strategy for executing the JOIN operation concerning data tables that are distributed and split to the two stores, and the redefinition of the query execution plan, in order for the query federator to exploit data locality and reduce the number of rows that will be retrieved and transferred from the object store via the network.
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Response time of the query execution is improved
Expected delivery date	M36 (3rd version of prototypes)

TABLE 212 - SOFTWARE REQUIREMENT REQ- SO-OR-05

Section	Description
ID	REQ- SO-OR-06
Title	Optimize access to Object Storage.
Level of detail	Software
Type	DATA (Data)
Description	To perform analytics efficiently on Object Storage, a client-side caching/acceleration layer is needed, which is critical for a hybrid cloud scenario, where some of the customer data is on premise (potentially LeanXscale database and Spark) and some in the cloud (potentially IBM COS). In such scenario, when performing analytics, data needs to move from COS to Spark across the WAN, minimizing the amount of data movement when part of the data is retrieved multiple times is of utmost importance. A similar scenario occurs in a multi-cloud environment, where data may be distributed among more than one cloud, requiring data transfer across the WAN
Additional Information	This complements data skipping and data layout techniques to further reduce the KPI measuring the number of bytes sent from Object Storage to Spark
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Response time of the query execution is improved
Expected delivery date	M36 (3rd version of prototypes)

TABLE 213 - SOFTWARE REQUIREMENT REQ- SO-OR-06

Section	Description
ID	REQ- SO-OR-07

Section	Description
Title	SQL Grammar extension
Level of detail	Software
Type	DATA (Data)
Description	In order to better support the seamless, an extension of the SQL grammar is needed
Additional Information	The grammar extensions will allow the database administrator to define that a data table can be split across the two datastores, and will allow him to provide additional information like the time window of the data slice, along with other configuration attributes like the minimum size of a data slice that is allowed to be moved, time frequency of the moving action etc.
Priority	DES (desirable requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Data user can use the standard JDBC connection with an extension of the SQL grammar to be able execute DDLs
Expected delivery date	M36 (3rd version of prototypes)

TABLE 214 - SOFTWARE REQUIREMENT REQ- SO-OR-07

Section	Description
ID	REQ- SO-OR-08
Title	Ensure data consistency when a moving action is taking place
Level of detail	Software
Type	DATA (Data)
Description	When data is moving from the operational store to the object store, data might either co-exist in both stores, or are non-existed in any store. The framework must be able to serve requests for data retrieval with no downtimes during this process, and the data should be consistent, meaning that the result of the execution of a query should be the same, no matter if the data are being moved
Additional Information	The operational datastore must not withdraw a data slice, until an acknowledgement of a persistence storage is being notified by the object store. In this case, data can co-exist in both stores. The Query Federator of the framework must take this into account, and re-write the queries to be executed in both stores accordingly in order to scan records on the visible data set in each store. In order to ensure data consistency when parallel transactions are being executed, before, during and after the data moving process, it will rely on the transactional manager of the operational datastore
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A

Success Criteria	Queries return equivalent results as before, during and after the moving of the data slice
Expected delivery date	M24 (2nd version of prototypes)

TABLE 215 - SOFTWARE REQUIREMENT REQ- SO-OR-08

Section	Description
ID	REQ- SO-OR-09
Title	Support JOIN operation between two fragmented tables
Level of detail	Software
Type	DATA (Data)
Description	The data user must be able to perform a JOIN operation over tables that are split across the two stores. Memory and network restrictions should be taken into account, along with the overall responsiveness
Additional Information	The execution of the JOIN operator requires the movement of the data from two tables. As the tables are distributed and located in different nodes (and database systems) this should minimize the amount of data that is being transferred over the network. Memory considerations must also be taken into account, as the query engine cannot retrieve all data (whose volume might be petabytes) in memory to execute the operation
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-04
Source	N/A
Success Criteria	Queries return equivalent results as before
Expected delivery date	M24 (2nd version of prototypes)

TABLE 216 - SOFTWARE REQUIREMENT REQ- SO-OR-09

Section	Description
ID	REQ- SO-OR-10
Title	Use optimization techniques to further improve the performance of the JOIN operator
Level of detail	Software
Type	DATA (Data)
Description	The use of bind join will reduce the number of tuples that need to be transmitted across the two data stores
Additional Information	For the hybrid part of the bind join, there is the need to overcome the limitation of Spark when submitting query statements involving an IN clause with a list of arguments of a very big size. This can be avoided by creating a temporal table in the object store, and execute the partial join there
Priority	OPT (optional requirement)
Reference Use Case	ALL
Role	ROL-04

Source	N/A
Success Criteria	Queries return equivalent results as before with lower response type
Expected delivery date	M36 (3rd version of prototypes)

TABLE 217 - SOFTWARE REQUIREMENT REQ- SO-OR-10

6.14 Policy Development Toolkit

Section	Description
ID	REQ- SO-PDT-01
Title	Store and retrieve information for policies and their relevant entities (i.e., KPIs, Stakeholders, Domains, Analytical tools, etc.)
Level of detail	Software
Type	FUNC (Function)
Description	The PDT Backend must store policies, KPIs related with the policies, their actors, the corresponding domains, the tools related with the KPIs and their results
Additional Information	Additional information can be found at the D5.3
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	Create a new policy and retrieve the information
Expected delivery date	M12 (1st version of prototypes)

TABLE 218 - SOFTWARE REQUIREMENT REQ-SO-PDT-01

Section	Description
ID	REQ- SO-PDT-02
Title	Provide meta-information of the policies and their related entities to the policy modelling editor to allow the latter to propose new policies
Level of detail	Software
Type	FUNC (Function)
Description	The PDT Backend must store additional meta-information related with the policies and relevant entities that can be consumed by the policy editor
Additional Information	Additional information can be found at the D5.3
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03

Source	N/A
Success Criteria	Policy Modelling Editor can propose valid policies based on the information that can be retrieved by the PDT
Expected delivery date	M24 (2nd version of prototypes)

TABLE 219 - SOFTWARE REQUIREMENT REQ-SO-PDT-02

Section	Description
ID	REQ- SO-PDT-03
Title	Retrieve list of available analytical tools
Level of detail	Software
Type	FUNC (Function)
Description	The PDT Backend must provide a catalogue of available analytical tools to the policy maker, so that he can be aware of what is supported by the tool
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to retrieve the list of available tools
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	Policy Maker can see the list of tools available, that must be the same with the tools already registered in the Data Analytic layer
Expected delivery date	M12 (1st version of prototypes)

TABLE 220 - SOFTWARE REQUIREMENT REQ-SO-PDT-03

Section	Description
ID	REQ- SO-PDT-04
Title	Invoke an analytical tool to perform an analysis
Level of detail	Software
Type	FUNC (Function)
Description	The PDT must be able to invoke the execution of an analysis from the tools available in the Data Analytic layer
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to request the execution of an analysis
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The analytical tool can be invoked by the PDT

Expected delivery date	M12 (1st version of prototypes)
-------------------------------	---------------------------------

TABLE 221 - SOFTWARE REQUIREMENT REQ-SO-PDT-04

Section	Description
ID	REQ- SO-PDT-05
Title	Get the status of pending analytical job
Level of detail	Software
Type	FUNC (Function)
Description	The PDT must be able to retrieve the status of a job that has been submitted for execution from an analytical tool
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to request the status of the job
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The status of the job can be retrieved by the PDT
Expected delivery date	M12 (1st version of prototypes)

TABLE 222 - SOFTWARE REQUIREMENT REQ-SO-PDT-05

Section	Description
ID	REQ- SO-PDT-06
Title	Retrieve the results of the analysis
Level of detail	Software
Type	FUNC (Function)
Description	The PDT must be able to retrieve the results of a job that has been submitted for execution from an analytical tool, so that can be visualized
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to retrieve the result of the job
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The result of the job can be retrieved by the PDT
Expected delivery date	M12 (1st version of prototypes)

TABLE 223 - SOFTWARE REQUIREMENT REQ-SO-PDT-06

Section	Description
---------	-------------

ID	REQ- SO-PDT-07
Title	Retrieve analytical jobs & their status/results submitted by a User
Level of detail	Software
Type	FUNC (Function)
Description	The PDT must be able to retrieve the history of analytical tools execution and their results that a User has submitted over time, along with meta-data (related to which policy and KPI)
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to retrieve and filter the related data
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The User transactions can be retrieved by the PDT and sorted by policy, KPI, date
Expected delivery date	M24 (2nd version of prototypes)

TABLE 224 - SOFTWARE REQUIREMENT REQ-SO-PDT-07

Section	Description
ID	REQ- SO-PDT-08
Title	Delete analytical jobs & results submitted by a User
Level of detail	Software
Type	FUNC (Function)
Description	The PDT must be able to remove user transactions / jobs and their results
Additional Information	A REST client must be implemented to communicate with the Data Analytic layer to remove jobs/status/results for a user invocation.
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The User transaction can be removed along with any job results
Expected delivery date	M24 (2nd version of prototypes)

TABLE 225 - SOFTWARE REQUIREMENT REQ-SO-PDT-08

Section	Description
ID	REQ- SO-PDT-09
Title	Forward analytics results to the Visualization component
Level of detail	Software

Type	FUNC (Function)
Description	Analytics results retrieved by the PDT backend must be further forwarded to the internal visualization component (REQ- SO-DV-01)
Additional Information	The format of the analytics results must follow the predefined protocol (in JSON), as to be compatible with the visualization functions
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The valid visualization component retrieves analytics results
Expected delivery date	M24 (2nd version of prototypes)

TABLE 226 - SOFTWARE REQUIREMENT REQ-SO-PDT-09

Section	Description
ID	REQ- SO-PDT-10
Title	Send data connectivity information to invoked analytical tools
Level of detail	Software
Type	FUNC (Function)
Description	The analytical tools should be transparent. The PDT should be able to send to them information regarding where they should connect to, in order to retrieve information
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The valid visualization component retrieves analytics results
Expected delivery date	M24 (2nd version of prototypes)

TABLE 227 - SOFTWARE REQUIREMENT REQ-SO-PDT-10

Section	Description
ID	REQ- SO-PDT-11
Title	The PDT must allow for analytical tools to combine information from various sources
Level of detail	Software
Type	FUNC (Function)
Description	The analytical tools might need to retrieve data from a combination of different datasets

Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03
Source	N/A
Success Criteria	The valid visualization component retrieves analytics results
Expected delivery date	M36 (3rd version of prototypes)

TABLE 228 - SOFTWARE REQUIREMENT REQ-SO-PDT-11

6.15 Data Visualization

Section	Description
ID	REQ- SO-DV-01
Title	Data Visualization
Level of detail	Software
Type	FUNC (function), L&F (Look and Feel Requirements), USE (Usability Requirements), PERF (Performance Requirements)
Description	End users must be able to check the data in a more visual way, through meaningful charts; that will help with insights and in taking decisions
Additional Information	N/A
Priority	MAN (mandatory requirement)
Reference Use Case	ALL
Role	ROL-03, ROL-04
Source	Data Analytics tools through the PDT
Success Criteria	Data Visualization component running and all required visualizations can be seen correctly
Expected delivery date	M24 (2nd version of prototypes)

TABLE 229 - SOFTWARE REQUIREMENT REQ-SO-DV-01

7 State of the Art analysis

This section presents the state-of-the-art analysis in the various sectors that the PolicyCLOUD project is being involved. Whenever is possible, it links the state-of-the-art technologies that are described in the following subsections with the context of the project and state how the platform can benefit from the use of those technologies.

7.1 Evidence based policy making and data analytics

Evidence Based Policy Making (EBPM) is a term usually applied when policy choices are performed based on objective evidences using a scientific approach, rather than based on intuition, random, ideology capricious, hidden interests or just theory. Even this approach is known since some centuries ago, it was the Blair administration in UK that brought it back to the political agenda in the late 90s to end “ideologically-based decision making and ‘questioning inherited ways of doing things’ [41]. EBPM can be used through all the policy making cycle [42], as seen in Figure 3: **1. Agenda setting**, to take decisions on which public issue requires the most attention to act; in **2. Policy formulation**, to define the strategies that can address the issue in the best way; in **3. Adoption**, to approve the regulatory measures based on objective advice; in **4. Implementation**, to implement the necessary infrastructure following a methodological approach that best supports the policy application; and **5. Monitoring and evaluation**, to assess if policies have reached their targets and therefore are successful or must be revised.

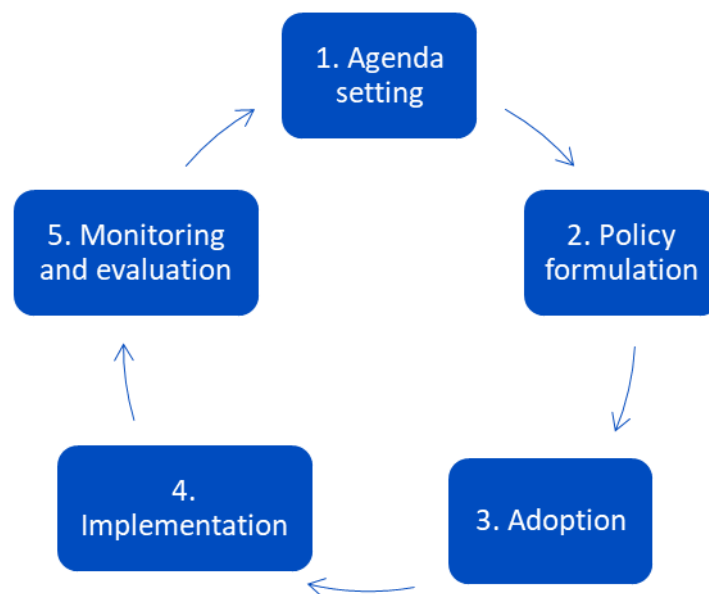


FIGURE 3: POLICY MAKING CYCLE

According to the Roadmap for future research directions from the Big Policy Canvas project [43], more advanced versions of this cycle exist, considering the use of big data analytics that enhance the previous cycle by providing evaluation capabilities to the whole cycle, and not only to the latest phase of monitoring and evaluation.

Many challenges are still not solved, as the use of data analytics in the frame of the policy making cycle is not a simple issue. It raises several challenges in the data gathering, integration and reuse. As well, the participation of several types of stakeholders raises many privacy and security issues. Additionally, there are problems in the use of artificial intelligence automation, for example, for biased decision-making because of the bias in training data used in this type of systems [43].

In addition to these issues, some bottlenecks, and enablers in the application of EBPM have been found [43]:

- Collection of big amounts of data is possible, but quality problems are still a big issue for the use of big data in public policies.
- Resources and budget limitations in public sector are often a burden that must be overcome.
- Interoperability issues with data from several sources, internal and external.
- Leadership issues and the impact of change of political direction after new administrations take over the government.
- Job market availability, limitations of data scientists.
- The importance of having a clear strategy and leadership for the use of data analytics results in the policy making process.
- Providing an opportunity for the update of legacy applications, improving efficiency and interoperability.
- Necessity to address applicable ethical, legal, regulatory, and societal requirements, as identified in D3.3 and D3.6.
- Improving the perception of efficiency of public sector, with high quality services with lower costs.
- Being careful about the use of big data technologies, as they may be a big opportunity for improving the service of public sector, but at the same time could be misused causing negative impacts to the citizens and even erode trust in public authorities.

Six research clusters, or open sets of questions, have been defined for the use of big data technologies in the scope of EBPM, according to Big Policy Canvas project [43], four of them according to the big data value chain and two which are horizontal to the whole value chain, as depicted in Figure 4.

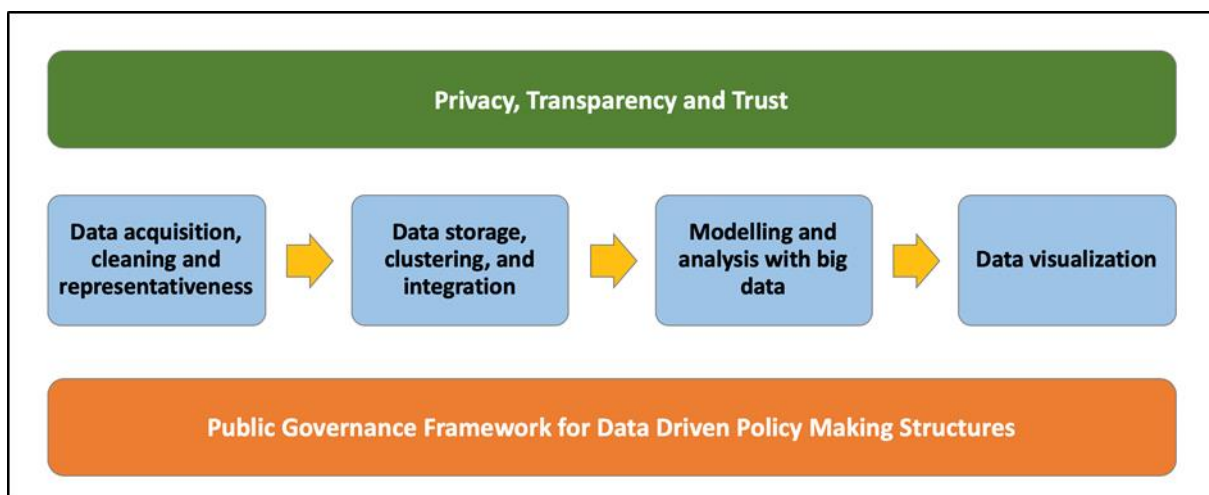


FIGURE 4 - RESEARCH CLUSTERS ACCORDING TO THE BIG POLICY CANVAS PROJECT

The horizontal **Privacy, Transparency and Trust** cluster deals, on one hand, with the issues related to data ownership, security and privacy, and, on the other hand, the issues related to the transparency of the policy making. For the security and privacy, for example, it requires tackling the issues derived from the data ownership and usage, ensuring that data collected is not used for other purposes than the ones for which it is provided (to this regard, a more detailed analysis is performed in D3.3 and D3.6). In this regard, research in the area of data regulation and standards for data generated by sensors, devices or in social media is still an open issue. As for the transparency in the policy making, the availability of public open data allows the scrutiny of these policy making processes and the policies implemented and their outcomes. To achieve this, the publication of public data under co-ownership conditions allows the audit and the reuse by society stakeholders. Additionally, algorithms for automated decision making and screening can lead to discrimination, traceability and lastly to breach data protection rules. To avoid this type of misuse, a set of ethical standards must be developed, ensuring the compliance during their design phase (to this regard, a more detailed analysis is performed in D3.3 and D3.6). Moreover, a potential area of research to ensure fairness in algorithms design is open in co-creation approaches for the design of public services when citizens are involved. Questions related to this cluster will be addressed by the Legal/Ethical Framework and privacy enforcement in the **PolicyCLOUD** architecture as part of **WP3 Cloud Infrastructures Utilisation & Data Governance**, and by the use-cases co-creation approach in **WP6 Use cases Adaptation, Integration & Experimentation**.

The second horizontal cluster, **Public Governance Framework for Data Driven Policy Making Structures**, refers to the set of rules to manage evidence-based policy-making to apply information technology in a way that it is possible to improve the policy-making process and better understand the underlying societal problems that have to be addressed. Implementation through the efficient use of data, achieving a rational, participative and transparent process is key. These aspects are addressed by some of the tasks in **WP5 Cross-sector Policy Lifecycle Management** of **PolicyCLOUD**, like, Cross-sector Policy Lifecycle Management for the modelling and design of policies, collection, experimentation, adaptation, optimisation and implementation of policies and their compliance monitoring.

The third cluster which is the first vertical one is the **Data Acquisition, Cleaning and Representatives**. This has to deal with the huge amount and variety of data sources, according to the origin (own public sector data, social networks, open data, private data), the types of data (internet of things, sensor data, real time data, geo-location, image, video, sound) and the sources (traffic and transport data, administrative processes, citizens, scientific). Meaningful conclusions are drawn up by extracting the behavioural essence associated to all this huge amount of data through the use of big data technologies. Nevertheless, there are some implicit issues with the use of such massive amounts of data in terms of quality, random and systematic errors. The impact of random errors can be minimised by the increase of the data size, but for systematic errors it is not the case. Anyhow, big data is not the Holy Grail that can provide an answer to all the questions, as it is prone to suffer problems due to bias and data inaccuracy, which can be avoided through cleansing mechanisms. These aspects are addressed on **PolicyCLOUD** in **Task 3.3 Cloud Gateways & APIs for efficient data utilisation**, and in **Task 4.2 Enhanced interoperability & Data Cleansing**; furthermore, a detailed analysis of the ethical, legal, regulatory, and societal risks related to big data has been performed in the context of D3.3.

The fourth cluster is the **Data Storage, Clustering, and Integration**. The use of data from so many and different sources, carries the implicit challenge of dealing with data that has not been produced with the goal of being collected for other specific uses. Therefore, the data collected is of heterogeneous nature, some is structured or semi-structured, and sometimes incomplete. For example, social media data is usually applied to sentiment analysis and opinion tracking processes, while it requires a lot of cleansing, and the resulting information is usually biased because of its nature. This is only one example of many data flows being continuously collected by all sorts of systems. Repurposing all this data in the policy-making domain requires data scientists and domain experts' skills to make the right interpretation of the data. In addition, it has to be considered the reuse of existing public sector data, and the availability of methodologies and infrastructures for the storage and processing of such big data. These issues are addressed in **PolicyCLOUD** by **Task 4.1 Cross-sector Data Fusion Linking**; also, a theoretical analysis of the ethical, legal, regulatory, and societal issues related to this cluster has been performed in the context of D3.3.

The fifth cluster is the **Modelling and Analysis with Big Data**. This deals with the approaches for modelling forecast scenarios based on big data, the data modelling and the simulation modelling, and the novel approaches and research being undertaken in this area. In **PolicyCLOUD** this is addressed by the tasks in **WP4** that deal with data analytics, **Task 4.3 Situational Knowledge Acquisition & Analysis**; **Task 4.4 Opinion Mining & Sentiment Analysis**; **Task 4.5 Social Dynamics & Behavioural Data Analytics**; and in **WP5** with **Task 5.2 Modelling & Design of Policies**.

Finally, the sixth cluster is the **Data Visualization**. The presentation of information from big data is a challenging issue, so the insights extracted from the data are presented in a meaningful way to humans. It is also relevant in the policy making context so to understand the problems from the results obtained from modelling and analytics tools. In this regard, the provision of evidence based on the identification of KPIs, and their relations is key. The most relevant approach for **PolicyCLOUD** data visualisation is the provision of dashboard visualisation to measure and monitor relevant indicators with respect to the final objectives for the corresponding policies. Other methods correspond to info-graphic presentations and

visual analytics. The visualisation approach is addressed by **Task 5.3 Policy Development Toolkit including Data Visualisation**; also, a theoretical analysis of the ethical, legal, regulatory, and societal issues related to this cluster has been performed in the context of D3.3.

7.2 Policy interoperable datasets

Nowadays, policy makers publish an increasing amount of their data on the Web in an effort with double fold meaning. In one hand, to comply with the emerging Open Data movement and in the other hand in order to optimize and improve their policy management and development lifecycle. A key to realizing the open data and providing advanced open policies is the ability to merge divergent data and datasets. Hence, interoperability is the key “back office” element across the whole policy making lifecycle and open data semantics [8]. Achieving true interoperability entails different representations, purposes, and syntaxes and will enable improved access to records, datasets and policies. Recent years many approaches, standards, ontologies and vocabularies have been proposed as means of achieving various tasks of interoperability between heterogenous and independent datasets. One of the first approaches on dataset interoperability is the Information Modelling and Interoperability (IMI) model, which further splits the interoperability into three distinct layers - the syntax layer, the object layer and the semantic layer [9]. Likewise, more recently the European Commission, through their program ISA² has defined the European Interoperability Framework (EIF) which defines interoperability across the above four layers - (i) organizational interoperability, (ii) semantic interoperability, (iii) technical interoperability and (iv) legal interoperability [10]. In addition, within LOD2 project the NIF framework was designed, which is based on a Linked Data enabled URI scheme for identifying elements in (hyper-)texts and an ontology for describing common semantic terms and concepts of NLP tools and services [11]. An emerging research direction entails automatically discovering links between datasets using Word Embeddings and other components that find links based on syntactic and semantic similarities [12]. Moreover, a recent research focused on implementing a vocabulary (i.e., VoDext) to formally describe virtual links in order to enable interoperability among different datasets [13]. By defining virtual links with VoDext RDF schema and by providing a set of SPARQL query templates to retrieve them, the research team achieved to facilitate the writing of federated queries and knowledge discovery among federated datasets. Furthermore, another project in the archaeological domain, also, highlights the use of RDF schemas to achieve dataset interoperability by extracting and exposing archaeological datasets (and thesauri) in a common RDF framework assisted by a semi-automatic custom mapping tool [14]. In addition, a relevant research introduced three metrics to express the interoperability between two datasets - the identifier interoperability, the relevance and the number of conflicts [15]. Another commonly used technology for achieving and enhancing interoperability is the JSON for Linking Data (JSON-LD) format, that has been a W3C recommendation since 2014 to promote interoperability among JSON-based web services [16]. A research in the biological sector highlights the usage of a JSON-LD system, which provides a standard way to add semantic context to the existing JSON data structure, for the purpose of enhancing the interoperability between APIs and data [17]. PolicyCLOUD project will enhance interoperability based on data driven-design, coupled with linked data technologies (e.g. JSON-LD and RDF) and standards-based ontologies and vocabularies to improve both semantic and syntactic interoperability. Moreover, a data

modelling by standard metadata schemas will be defined in order to specify the metadata elements that should accompany a dataset within a domain. To this end, linked data will work as the foundation of a common export format for data within PolicyCLOUD Marketplace.

7.3 Enhanced visualizations providing actionable insights

Visualizations are the most understandable way for humans to show the results of data analysis. Well known is the saying - "An image is worth a thousand words", and following this idea, through different kinds of visualizations, images allow heterogeneous users to obtain in a concise, ordered and structured way, a broader knowledge of the information they need in each time, and, consequently, a better decision making. In this line, charts, for example, are a kind of visualization that is widely used because of its easy representation and users' understanding.

In order to visualize data as a chart in a web site there are three different approaches:

- **Use products/tools/software created for that use:**

In the case of products, there are some software tools in the market that have been designed to visualize data according to the most common needs of companies. Some of these products are:

- Datapine¹⁰ - A Software as a Service (SaaS) platform that can be used to display data as charts.
 - Microsoft Excel - Microsoft excel is a spreadsheet software developed by Microsoft where there is the possibility to create several different types of charts.
 - Grafana¹¹ - An open source analytics solution that allows us to visualize data in order to understand the trends of it.
 - Tableau¹² - Offers a platform to display charts.
 - Microsoft PowerBI¹³ - Part of the Microsoft Office 365 package; it offers several charts to display the information of a company.
 - QlikView¹⁴ - An End-to-End Data integration and analytics tool.
- **Create charts into the backend of the site as image and display them into the front-end:**

Backend libraries are used by the backend of web sites to create charts when it receives the petition. Some of these libraries are:

- JpGraph¹⁵ - A PHP library, valid for PHP5 and PHP7 that can create several types of charts.

¹⁰ <https://www.datapine.com/>

¹¹ <https://grafana.com/>

¹² <https://www.tableau.com/>

¹³ <https://powerbi.microsoft.com/>

¹⁴ <https://www.qlik.com/>

¹⁵ <https://jpgraph.net/>

- Matplotlib¹⁶ - A Python 3 library that can be used to create static, animated, and interactive visualizations.
- **Create dynamic charts in the front-end on the fly using JavaScript libraries:**

JavaScript Libraries to be used in a front-end.

In the scope of the project, the chosen way to do it is to create a variety of dynamic charts in the front-end, on-the-fly, when end users wish to display the data. This way is lighter for the server and the results are more attractive, useful and understandable for users, mostly because of its interactive feature. The chosen language is JavaScript (JS), the most used language for webs. Following this approach there are several JavaScript libraries that can be used to create charts. These libraries use one of these rendering technologies - HTML5 Canvas, SVG (Scalable Vector Graphics) or VML (Vector Markup Language) to create charts. This technique usually needs to request data from the backend by APIs that return data in JSON formats.

Some of the most popular chat libraries are shown in the table below:

Library name	Main Site	License	Rendering technology	Public repository	code
amCharts	https://www.amcharts.com/	Proprietary	SVG and VML	https://github.com/amcharts/amcharts4	
AnyChart	https://www.anychart.com/	Proprietary	SVG and VML	https://github.com/AnyChart/AnyChart	
C3.js	https://c3js.org/	MIT	SVG	https://github.com/c3js/c3	
Chartist.js	https://gionkunz.github.io/chartist-js/	WTFPL or MIT	SVG	https://github.com/gionkunz/chartist-js	
Chart.js	https://www.chartjs.org/	MIT	Canvas	https://github.com/chartjs/Chart.js	
D3.js	https://d3js.org/	BSD-3	SVG	https://github.com/d3/d3	
Flot	https://www.flotcharts.org/	MIT	Canvas	https://github.com/flot/flot	
FusionCharts	https://www.fusioncharts.com/	Proprietary	SVG and VML	https://github.com/fusioncharts/	
Google Charts tools	https://developers.google.com/chart	Free	Canvas, SVG and VML		

¹⁶ <https://matplotlib.org/>

Highcharts	https://www.highcharts.com/	Proprietary	SVG and VML	https://github.com/highcharts
Plotly.js	https://plot.ly/javascript/	MIT	SVG	https://github.com/plotly/plotly.js
Ngx-charts	https://swimlane.github.io/ngx-charts/#/ngx-charts/bar-vertical	MIT	SVG	https://github.com/swimlane/ngx-charts

TABLE 230 - MOST POPULAR JS CHART LIBRARIES [19], [20], [21], [22], [23]

When developing visualizations, using a framework usually helps to save time and efforts, as they can facilitate the whole process. JavaScript frameworks make it easier developing with JavaScript. Some of the most popular JavaScript frameworks are shown in the table below:

Framework	Main Site	License	Current version	Size
React	https://angular.io/	MIT	v9.1.0	143K
Vue	https://reactjs.org/	MIT	v16.13.1	43K
Backbone	https://vuejs.org/	MIT	v2.6.11	23K
Ember	https://backbonejs.org/	MIT	v1.4.0	7.3K
Meteor	https://emberjs.com/	MIT	v3.17	95K
Nodejs	https://www.meteor.com/	MIT	v12.16.1	N/A
Mithril	https://nodejs.org/	MIT	v13	N/A
Polymer	https://mithril.js.org/	MIT	v2.2.0	N/A

TABLE 231 - MOST POPULAR JS FRAMEWORKS [24], [25]

In the table below it is shown which libraries from above table, can be integrated easily in with each framework:

	Ang ular	Reac t	Vue	Back bon e	Emb er	Met eor	Nod ejs	Mith ril	Poly mer	Tota l
amCharts	X	X	X		X	X	X		X	7
AnyChart	X	X	X		X	X	X			6
C3.js	X	X	X	X	X	X	X	X	X	9
Chartist.js	X	X	X			X	X		X	6
Chart.js	X	X	X	X	X	X	X	X	X	9
D3.js	X	X	X	X	X	X	X	X	X	9
Flot	X	X	X		X	X	X		X	7
FusionCharts	X	X	X	X	X		X		X	7
Google Charts tools	X	X	X	X	X	X	X		X	8
Highcharts	X	X	X		X	X	X		X	7
Plotly.js	X	X	X		X		X			5

ZingChart	X	X	X	X	X		X			6
Ngx-charts	X									
Total	13	12	12	6	11	9	12	3	9	

TABLE 232 - COMPATIBILITY BETWEEN JS LIBRARIES AND FRAMEWORKS

As can be seen in the previous table, four frameworks have more JavaScript chart libraries compatibilities - AngularJS, React, Vue and Nodejs. The final decision about which framework and which JavaScript library will be used will not take only this information into account, but also if there are any other project requirement that directly affect this decision.

7.4 Re-usability of analytical tools/models decoupled from infrastructure

It is clear today that the modern cloud native architecture should be based on the micro-services pattern¹⁷ where the application is decomposed into modular independent components which consume each other's API to provide the overall functionality. The definite major winning technology today for achieving this decomposition is containers. In addition to the decomposition of application components, containers decouple the application from the underlying infrastructure. Packaging within containers ('Containerization') provides dramatic simplification and speed of deployment on any cloud infrastructure, as well as avoidance of lock-in to any platform. Recoverability, elasticity and scalability properties of the underlying platform are additional benefits of the decoupled architecture. For containers, the dominate core container technology is Docker¹⁸ and the container framework technology is Kubernetes¹⁹.

For analytical tools, specifically over big data, the advances in data centre networking make the disaggregation of storage and compute the widely employed pattern today. The decoupling of the storage infrastructure provides even greater advantage over the general application decomposition due the complexity and price of storage platforms, having the ability to re-use and apply any analytical tool on big data residing on any storage platform. An excellent example is Apache Spark²⁰, today the most popular open source for big data analytics that supports various analytic frameworks as GraphX for graph processing, MLlib for machine learning, SQL and streaming. Each of its analytic frameworks can work with any storage platform (as HDFS, S3 Object Storage) decoupled from the computing resources.

¹⁷ <https://www.ibm.com/cloud/learn/microservices>

¹⁸ <https://www.docker.com>

¹⁹ <https://kubernetes.io>

²⁰ <https://spark.apache.org>

This enables to decouple the analytic logic and modelling from the cumbersome setting and integration details for each platform.

Emerging specialized data format frameworks provide today intermediate layer for the analytics' logic and modelling, exploiting seamlessly features of underlying storage platforms. The major ones today are Delta Lake²¹, Apache Iceberg²² and Apache Hudi. These frameworks provide abstraction over table/file formats for analytics tools, with consistency and performance optimization features as data catalog / meta data, schema and layout evolution, time travel, atomicity, Merge on Read, Copy on Write, limited transactional operations and exploitation of columnar formats as Parquet²³. A layer above the data format frameworks is the data warehouse / data lake frameworks which provide consistent and controlled access to various data sets and data sources, as Snowflake²⁴ which hide the actual cloud infrastructure (runs on Amazon, MS Azure, Google Cloud Platform) and Dremio^{25 26} which provides optimized data lake engine based on Apache Arrow for in-memory columnar data processing. These abstraction layers enable data scientists to concentrate on the analytic algorithms and models, and reuse them over different compute and storage infrastructures in much greater ease than in the past, although of course the picture is not perfect and migration from one platform to another is usually not completely transparent.

7.5 Polyglot analytical tools federating heterogeneous sources and stores

Accessing heterogeneous data sources (a concept often addressed by data integration systems or multidatabases [36], [39] is a problem that has been widely studied in the literature and with the recent emerge of cloud databases and big data processing, it has been evolved towards polystore systems. The latter provide a common accessibility method in order to retrieve data from a variety of heterogeneous target data stores, such as typical relational DBMS, NoSQL or NewSQL datastores, or HDFS datalakes, involving data that can be either structure, semi-structure or fully unstructured. Their early implementations [34], [35], [38] relied on a single common model that the target datastores had to transform their schema to. A further improved presented by the polystore BigDAWG [31], [32] which defines *islands of information*, where each island is related to a specific data model and language in order to provide access to the underlying data store. It additionally provides the support for queries spanning among the different data models by moving the intermediate datasets between those islands.

²¹ <https://delta.io>

²² <https://iceberg.apache.org>

²³ <https://parquet.apache.org>

²⁴ <http://pages.cs.wisc.edu/~remzi/Courses/739/Spring2004/Papers/p215-dageville-snowflake.pdf>

²⁵ <https://www.dremio.com>

²⁶ <https://github.com/dremio/dremio-oss>

Moreover, Myria [40] uses a shard-nothing parallel architecture for data federation across the heterogeneous models and query languages and exploits its extended relational model and its unique imperative language for defining transformation rules that will allow the input query to be applied to a target datastore-specific call. It is worth to mention that other polystore solutions [29], [30], [33] rely on the application requirements themselves to decide the optimal data placement and the query execution plan.

Spark SQL [28] is a parallel SQL engine that offers tight integration between traditional relational and procedural processing via a standard API, taking advantage of massive parallelism. It offers a DataFrame API that translates relations into arbitrary object collections, thus supporting operations targeting external datastores that transform the data into those collections. It makes use of data connectors implemented for each supported datastore whose role is to map a data item into this DataFrame. Presto [37] on the other hand is a distributed SQL query engine which makes use of interactive analytic queries against the target datastores. When it comes to the query execution, it allows for massively parallel processing, consisting of a coordinator and multiple workers, each one of those is making use of target specific connectors which implements a common interface. The implementation of the latter encapsulates the target database details on how to access the data source, while provides the data schema metadata to the coordinator to be taken into account during the query plan of the execution. What is more, Apache Drill [26] is another distributed query engine for large-scale datasets that is capable of querying data coming from various data sources via their own plugins implemented for each one of the latter. It also uses massively parallel processing that allows for scaling to thousands of nodes while maintaining overall latency, even if when processing petabytes of data. Each of its workers, called *drillbits* in its terminology, receives a query and compiles it accordingly and decides over an optimized query execution plan that can be parallelized taken into account data locality. Finally, Impala [27], which also provides a massively parallel processing engine, ensures overall low latency and high concurrency for analytical queries, making use of data specific connectors that transform the retrieved data from an external dataset into Hadoop compliant format, and then makes use of MapReduce jobs combining the intermediate results.

All of the aforementioned solutions that can be considered as polyglot analytical tools that enable the federation of heterogeneous datastores relies on their own specific data model and query language for query execution, and provide technology-specific interfaces for the integration with the applications and the data user analytical tools. PolicyCLOUD will rely on the engine of its central repository provided by LXS, which not only enables for query parallelism with the external datastores, but also supports the combination of massive parallelism with native queries and the optimizability of join operations, which is addressed by the LeanXcale distributed query engine.

7.6 Efficient data fusion from various data sources

For PolicyCloud scenarios, two aspects of efficient data fusion are important – (1) scalability of massive data ingest from multiple data sources, where burst of incoming metric data may lead to analytic results of required urgent alert in respect to some policy validation rule, and (2) the capability to apply analytics

over incoming data in flight, with the intent of storing only the resulted insight rather than the whole bulk data.

For the first aspect, scalable data fusion frameworks are mostly deployed for IoT scenarios as smart cities, where incoming data from multiple sensors or other IoT devices needs to be efficiently processed and stored. These frameworks are classified by several categories [1]:

- **Objectives** - Fixing problematic data [3], Improving data reliability, Increasing data completeness; ensuring the implementation of compliance requirements related to data sources
- **Techniques** - Data association – correlation between sources, Increased state estimation by inspecting multiple sources, Prediction, Unsupervised ML, Dimension reduction – for feature extraction e.g. in PCA
- **Data Input and output types** - Data2data, Data2feature, Feature2feature
- **Data source types** - Physical – as temperature or air quality sensors, Cyber – internet sources as web access data and social network data, Participatory – crowdsourcing from data contributed by personal devices [5], Hybrid – data obtained from mixed types of sources [6]
- **Scale** - Sensor level [4], Building wide [7], Inter-buildings, City wide, Inter-city
- **Platform Architectures** - Edge - data sources are processed at the edge, Fog - data sources are processed at a middle layer between the edge and the cloud, e.g. at cloud gateway, Cloud - data sources are processed in the cloud, this is the most common technique practiced by industry and research institutes for processing big data, Hybrid - processing is done in two or more layers (edge, fog and cloud) [2]

For the second aspect, i.e. applying analytics over incoming data in flight with the intent of storing only the resulted insight, there are several tools, most of them are open source or have open source version. One of the major commonly used tools today that enables analytics on streaming data is Apache Spark Streaming²⁷ which enables to apply core Spark analytics within live stream processing. It supports various streaming sources as Kafka, Flume, Kinesis, TCP sockets, and the processed data can be pushed to filesystems, databases, or dashboards. Internally Spark Streaming divides the incoming streams into batches, which then can be processed by regular Spark and generate stream of batch results. Another emerging streaming processing tool is KSQL²⁸ which is open source and Confluent KSQL²⁹ which is extended commercial version. It provides SQL interface for stream processing above Apache Kafka³⁰, and even the open source version is designed for mission-critical and scalable deployments. It provides a very simple programming interface (relative to Spark), and supports numerous streaming operations, including

²⁷ <https://spark.apache.org/streaming>

²⁸ <https://github.com/confluentinc/ksql>

²⁹ <https://www.confluent.io/product/ksql>

³⁰ <https://kafka.apache.org>

data filtering, transformations, aggregations, joins, etc. Other tools are Flume³¹ which is supported in many commercial Hadoop distributions, Apache NIFI³² that is used for data processing among multiple sources and targets, Apache Storm³³ that is used in many real-time deployments, and Amazon Kinesis³⁴ for real-time data analytics in the AWS eco-system.

7.7 Efficient cloud infrastructures

The revolution in information technologies we are facing over the past decades has played a decisive and unprecedented role in the development of society, science, technology, and economics. Today we are living in the big data era. Data volume is continuously increasing, doubling every 3 years. Within one minute, 400 hours of videos are uploaded on YouTube, 3.6 million Google searches are conducted worldwide each minute of every day, more than 656 million tweets are shared on Twitter, and more than 6.5 million pictures are shared on Instagram each day. When a dataset becomes so large that its storage and processing become challenging due to the constraints of existing tools and resources, the dataset is referred to as big data. When dealing with huge data volumes to be analyzed, cloud compute and big data come to play as they provide a solution which is both scalable and accommodating for big data analytics. Through hardware virtualization, cloud computing provides the option of storing significant amounts of data with the help of scalability, fault tolerance and availability. This allows Big Data to be available, scalable and fault tolerant through cloud computing. From a technical perspective, cloud computing consists of three **service models**, which can be offered across three **different deployment models**. The service models consist of Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS):

- Software as a Service, also known as cloud application services, represents the most commonly utilized option for businesses in the cloud market. SaaS consists of firms offering the capability to use software applications that are housed off the user's premises. SaaS utilizes the internet to deliver applications, which are managed by a third-party vendor, to its users. A majority of SaaS applications run directly through web browsers, which means they do not require any downloads or installations on the client side.
- Cloud platform services, also known as Platform as a Service (PaaS), allow users to develop their own Web-based applications or to customize existing applications using one or more programming languages and development tools. These services can be considered natural extensions of individual computer platforms. All servers, storage, and networking can be

³¹ <https://flume.apache.org>

³² <https://nifi.apache.org>

³³ <http://storm.apache.org>

³⁴ <https://aws.amazon.com/kinesis>

managed by the enterprise or a third-party provider while the developers can maintain management of the applications.

- Cloud infrastructure services, also known as Infrastructure as a Service (IaaS), allow customers to access the equipment and hardware needed to perform computing operations, including storage, processing and networking components.

The three deployment models through which these services can be provided, separately or together, are - (1) private clouds, (2) public or 'community' clouds, or (3) hybrid structures. Private clouds are exclusive to a single user. Public or community clouds are available to the general public or shared by large diverse groups of customers, and hybrid clouds combine public and private elements in the same data center.

For PolicyCLOUD a federated IaaS cloud infrastructure will be set-up for helping policy makers, public authorities, and different stakeholders to analyze a wide plethora of datasets collected from different data sources and provide decision making support to public authorities for policy modelling, implementation, and simulation, as well as for policy enforcement and adaptation. Further details about the federated IaaS-type cloud infrastructure used during the project have been provided in D2.6.

7.8 Data Governance Model, Protection and Privacy Enforcement

The data governance model and the tools for protection and privacy enforcement will be used to protect data and ensure decisions across the complete path follow specific guidelines and legislations. Although we are currently not able to precisely define the data governance model and its attributes, we have examined the relevant efforts on models, standards and frameworks for achieving our objectives.

The Data Governance Model will be based on the usage of the Responsibility assignment matrix (RACI) model and will be used to ensure access of different entities to the corresponding datasets at specific phases of the data and the policy lifecycles. RACI is used for clarifying and defining roles and responsibilities in cross-functional or departmental projects or business processes, and it builds on four key responsibilities most typically used. The responsibilities used are the following;

- **Responsible** refers to the people who do the work to complete the task.
- **Accountable** is a single person specified answerable for the correct and thorough completion of the deliverable or task and approves work that responsible provides.
- **Consulted** refers to people whose opinions are taken into account and with whom there is a two-way communication.
- **Informed** refers to people that kept up-to-date on progress, often only on completion of the task or deliverable; and with whom there is just one-way communication.

RACI model defines roles and people - a role is a descriptor of an associated set of tasks; may be performed by many people, and one person can perform many roles.

For the PolicyCLOUD Data Governance Model, we have to adapt RACI to be able to model the access of specific stakeholders to specific data at specific points in the lifecycle. For achieving this we have to use an Access Control Mechanism, and in specific Attribute-based access control (ABAC). Access Control Mechanisms are mechanisms realizing various logical access control models that provide the framework and set of boundary conditions upon which the objects, subjects, operations, and rules may be combined to generate and enforce an access control decision.

Several models and mechanisms, with each having its own advantages and limitations. For the sake of completeness, the most dominant ACMs will be listed:

- **Discretionary Access Control (DAC)** where the owner of the object specifies which subjects can access the object. Most operating systems such as all Windows, Linux, and Macintosh and most flavors of Unix are based on DAC models.
- **Mandatory Access Control (MAC)** where the system (and not the users) specifies which subjects can access specific data objects. The MAC model is based on security labels. Subjects are given a security clearance (secret, top-secret, confidential, etc.), and data objects are given a security classification (secret, top-secret, confidential, etc.).
- **Identity Based Access Control (IBAC)** uses mechanisms such as access control lists (ACLs) to capture the identities of those allowed to access an object. In the IBAC model, the authorization decisions are made statically prior to any specific access request and result in the subject being added to the ACL.
- **Role-based access control (RBAC)** employs pre-defined roles that carry a specific set of privileges associated with them and to which subjects are assigned.
- **Attribute-based access control (ABAC)** uses attributes, and policies that express boolean rule sets that can evaluate many different attributes before allowing access. ABAC, therefore, avoids the need for capabilities (operation/object pairs) to be directly assigned to subject requesters or to their roles or groups before the request is made. IBAC and RBAC can be seen as special cases of ABAC, with IBAC using the attribute of “identity” and RBAC using the attribute of “role”.

The adaptability and expressiveness of ABAC make it ideal for protecting the data in the lifecycle of PolicyCloud. The key standards that implement ABAC are OASIS standard of extensible Access Control Markup Language (XACML)³⁵ and the Abbreviated Language For Authorization (ALFA). XACML uses XSD notation in order to model the three basic artefacts (policy, the request and the response) which are required in an authorization scenario. ALFA is a pseudocode language that respects the XACML model (contains the same structural elements as XACML i.e. PolicySet, Policy, and Rule), but uses JSON instead

³⁵ <http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html>

of XML for the definition of access-control policies and maps directly into XACML. More information about ABAC and XACML will be provided in the deliverables D3.1/D3.4/D3.7.

For the PolicyCloud Data Governance Model, we have to identify a set of properties (as part of the ABAC concept) regarding the data, the data sources/origins, the phase of the data lifecycle (e.g. stored data or analysed data) and the phase of the policy lifecycle (e.g. modelling or experimentation process).

Finally, regarding the actual implementation of the privacy enforcement mechanism, the ABAC based authorization should be performed through the evaluation of policies per each data access request. Different tools will be examined in order to select, adapt and extend the most appropriate for PolicyCloud, namely PaaSword³⁶, Drools³⁷, Keycloak³⁸ and WSo2 Balana³⁹, AuthzForce⁴⁰, based on the trade-off between flexibility, expressivity support and efficiency.

7.9 Cross-sector Policy Lifecycle Management

While several application domains are exploiting the added-value of analytics over various datasets to obtain actionable insights and drive decision making, the public policy management domain has not yet taken advantage of the full potential of the aforementioned analytics and data models. Diverse and heterogeneous datasets are being generated from various sources, which could be utilized across the complete policies lifecycle (i.e. modelling, creation, evaluation and optimization) to realize efficient policy management. Although it is imperative that policymaking is based on scientific evidence, in many countries, particularly low- and middle-income countries, evidence-informed decision-making remains the exception rather than the rule [44]. Even into high-income countries internal data, reports and the opinions of internal staff members are the kinds of information used most frequently instead of research evidence [45], [46].

Agent-based dynamic simulation platforms to identify beneficial policies and interventions have been recently reported for cases such as the impact of sugar-sweetened beverage warning labels [47], the relation of urban crime with obesity [48], and reducing alcohol-related harms [49]. The platforms take residential and sociodemographic data and, through experimental scenarios, estimate the probability and the evolution of various factors. The participation of stakeholders in running simulations and different scenarios builds an [50].

³⁶ <https://paasword.io/>

³⁷ <https://www.drools.org/>

³⁸ <https://www.keycloak.org/>

³⁹ <https://github.com/wso2/balana>

⁴⁰ <https://authzforce.ow2.org/>

Recently, many projects have been spawned in the direction of evidence-based policymaking via the effective use of big data analytics. We can divide them into two categories. Into the first category, heterogeneous big-data datasets are collected, even real-time, to produce quantitative evidence supported by what-if scenarios. Such projects are BigO⁴¹ (Big data against childhood obesity), and MIDAS⁴² (Meaningful Integration of Data, Analytics and Services).

In the second category there is an additional layer, where the scientific evidence is framed in a way to support the formulation of public policy models and their management. The EVOTION⁴³ Project (Big data for hearing loss interventions), is one of the few attempts with specific outcomes, to formulate evidence-based policies. PHP decision making (PHPDM) models are structures having the following set of building elements - Goals, Objectives, Decision Criteria, Data, Factors, Types of Analysis and Policy Actions [51]. The ontology instance of the PHPDM is compiled through a reasoner, producing the corresponding Big Data Analytics (BDAs) components for the delivery of quantitative results⁴⁴. CrowdHEALTH⁴⁵ is an international research project co-funded by the European Commission that integrates high volumes of health-related heterogeneous data from multiple sources with the aim of supporting policymaking decisions [52]. The front-end of the platform is a health policy creation and evaluation environment, which provides advanced decision support, through data-driven analytic tools, both in aggregate as well as in personalized fashion. It presents a modular architecture and a secure big data processing workflow [53], while the Public Health Policy Model (PHPM) structure has elements consisting of Actors, Stakeholders, Key Performance Indicators KPIs, Formula (for the computation of the KPIs, Data, and Health Analytics Tools [54].

From the reported early attempts to develop platforms assisting policymakers to benchmark, simulate and forecast outcomes of policy decisions, we can discern challenges towards many directions, some of which are listed here:

- Representing a policy with measurable and quantitative variables.
- Finding, collecting, converting and handling big data sources at spanning time scales.
- Covering sensitivity of personal data, security and trustworthiness.
- Distributed reusable Big Data Analytics independent of cloud vendors, architectures or analytics frameworks.
- Full tracking and versioning of developed Policy Models along with supporting evidence and confidence intervals/error metrics.
- Hiding technical complexity, providing easy interaction of the policymaker with the platform.

⁴¹ <https://bigoprogram.eu/>

⁴² <http://www.midasproject.eu/>

⁴³ <http://h2020evotion.eu/>

⁴⁴ <https://scite.ai/reports/towards-a-model-driven-platform-for-6Pm3lr>

⁴⁵ <https://crowdhealth.eu>

To address these challenges, along with the complexity of the policy formulation and lifecycle management, there are many ICT tools spanning across categories - Visualization tools, Argumentation tools, eParticipation tools, Opinion mining tools, Incentive Management Tools, Simulation tools, Serious games, Persuasive tools, Social network analysis (SNA) tools, Big data analytics tools, and Semantics & linked data tools [55].

In a recent work, Giabbanelli et al. [56] highlight five areas where policy-making supporting software should assist -

- Participants access and update supporting definitions and evidence.
- Manually exchange of information between the new software and visualization, argumentation, and simulation tools.
- Iterative process to discern policy ‘inputs’ within context (loops in cognitive maps).
- Ability to monitor the outcomes of interventions via disjoint paths.
- Finding and filtering rippling effects of interventions.

Summarizing, the latest developments in big data analytics and the vast amounts of data that are being generated by different sources provide an opportunity for optimizing cross-sector policy lifecycle management, enabling public authorities and stakeholders to create, analyse, evaluate and optimize policies based on the “fresh” data, the information that can be continuously collected by citizens and other sensors. These technologies will be further examined and exploited within the framework of PolicyCLOUD so as to provide an integrated web-based environment to fulfil the requirements of advanced policy lifecycle management.

8 Background Technologies

The development of the PolicyCLOUD platform has been based on already existed baseline technologies that the partners of the consortium are bringing to the project as the background and they plan to further develop them in order to fulfil the requirements of the platform, as they have been listed in the previous subsections. The following table contains a list of those indicative baseline technologies that had been considered as candidate to be exploited by the PolicyCLOUD platform.

Technology Name	Technology Description	Advancements / Usage
LeanXcale DataStore	A highly scalable relational database management system, that ensures transactional semantics and provides an efficient manner to deal with highly ingestion rates. Additionally, it offers an parallel analytical query engine that makes it possible to retrieve data, while data is being ingested, thus, can be considered as an HTAP database, which allows for combining analytical and operational workload on the same data. It also can be easily extended to provide polyglot support.	In the scope of PolicyCLOUD, the internal query engine of the datastore will be extended in order to achieve a greater level of parallelism, and thus being able to serve analytical queries much more efficiently. Moreover, its internal polyglot support will be further extended in order to provide a common manner to access and join data storing in other stores, contributing to the fusion of the data.
Capturean Tool	Solution for SN monitoring. Available for Twitter mostly. Provides sentiment analysis and other SN metrics for specific listening channels (topics) defined by the customers. The solution provides a dashboard to visualize the results and a REST API to access to it programmatically	Sentiment Analysis over RRSS (mainly Twitter)
IBM Data Skipping and Smart Layout library	Library for Apache Spark for creating and using metadata indices that optimize SQL-based analytics, adjust the data layout for analytics optimization.	Dramatic performance improvement for SQL-based analytics over big data in object storage
Policy Development Toolkit	A framework for creating and evaluating policies related with the healthcare section	In the scope of PolicyCLOUD this asset will be further extended in order to allow the creation of general scope policies
Interoperability mechanism	Data interoperability realized through a process of identifying the structural and semantic similarity of domain-specific knowledge to turn the datasets into interoperable domain-agnostic ones.	In PolicyCLOUD the data interoperability mechanism will be extended towards different types of datasets and formats emerging from the identified PolicyCLOUD underlying data sources

Sources reliability tool	Sources reliability tool for mapping heterogeneous IoT devices into specific levels of trustfulness, thus estimating the overall reliability of each data source.	In PolicyCLOUD the sources reliability tool will be extended for estimating the reliability of all the available different types of data sources, and thus keeping into the platform for further analysis only the data that comes from only reliable sources.
--------------------------	---	--

TABLE 233 - BASELINE TECHNOLOGIES

Moreover, the majority of the partners of the consortium have great experience in participating in other on-going European and National research projects, whose topic of interest are relevant to the PolicyCLOUD. Due to this, outcomes and assets that have been developed in those research projects are candidates to be part of the platform or to further extend their functionalities in order to address PolicyCLOUD specific requirements. The following table contains a non-exhaustive list of other projects that might be useful in the development of the platform, along with information on how the latter can benefit from their use.

Project	Relevant Result	Advancement in PolicyCLOUD
CrowdHEALTH	<ul style="list-style-type: none"> We will use the Policy Development Toolkit that was firstly introduced in the scope of the CrowdHEALTH project. This is a framework for Analytics Tools registration into the Back-end and communication with the Policy Development Toolkit front-end Extended model of interoperable health data, and mechanism for estimating data sources' reliability 	<ul style="list-style-type: none"> UI Dashboard for policy development. Personal workspace for policy makers. Parameter selection functionality for the invocation of Analytics. User notification for Analytics results. On-line help provision to the policymakers Utilize the interoperability model, and the data sources reliability calculation metrics for additional datasets/cases identified in the context of PolicyCLOUD
CoherentPaaS	<ul style="list-style-type: none"> A common query language that can be used from a polystore in order to retrieve data resides in different and heterogeneous datastores. A polyglot query engine that can execute queries addressing different datastores 	<ul style="list-style-type: none"> The outcomes of CoherentPaaS will be used in order to implement the data fusion of the platform. They will be further extended to achieve the maturity required for the needs of the project, as the current state is a prototype. Further extensions will be made to support the different use cases
BigDataStack	<ul style="list-style-type: none"> Data layout extensions for the IBM Object Store, in order to accelerate analytical queries. A seamless analytical framework that combines the 	<ul style="list-style-type: none"> More accurate data layout. Support for all SQL data operations from the seamless analytical framework

	benefits of an operational database and data warehouse, moving historical data from the former to the latter	
DECIDO	<ul style="list-style-type: none"> demonstrate the ground-breaking impact of the adoption of innovative methodologies, tools and data enabling the effective development of better evidence-based policies by public authorities 	<ul style="list-style-type: none"> Sharing of best-practices/expertise and solutions to support the needs of public authorities
AI4PublicPolicy	<ul style="list-style-type: none"> a joint effort of policymakers and Cloud/AI experts to unveil AI's potential for automated, transparent and citizen centric development of public policies. To this end, the project will deliver, validate, demonstrate and promote a novel Open Cloud platform for automated, scalable, transparent and citizen-centric policy management based on unique AI technologies 	<ul style="list-style-type: none"> Sharing of best-practices/expertise and solutions to support the needs of public authorities
InteropEHRate	<ul style="list-style-type: none"> Research data sharing protocol (RDS) for health data exchange between citizens and researchers (upon consent approval) for further analytical and research purposes 	<ul style="list-style-type: none"> Usage of PolicyCLOUD's analytical tools on collected data towards the creation of policies in the healthcare domain
NloVe	<ul style="list-style-type: none"> NloVe project is making a step-change in Europe in bringing down the number of fatalities, reducing harmful emissions from transport work and reducing congestion within urban environments, by exploiting IoT technologies. The aim of nloVe is to detect cyber-attacks in real-time while simultaneously preventing them 	<ul style="list-style-type: none"> The expertise related to NloVe will be used to address all the relevant ethical, legal, regulatory, and societal issues related to PolicyCLOUD. Furthermore, the specific expertise related to transportation will be particularly useful with regards to the third use case

TABLE 234 - RELEVANT RESEARCH PROJECTS

9 Conclusion

This document firstly summarizes the methodology that was agreed in the scope of 'T2.1 Requirements Elicitation & State of the Art Analysis' of the project for collecting the user and technical requirements of the project. Based on this methodology, the list of concrete scenarios for each of the use cases has been extended, as well as the initial version of their relevant user requirements. An updated set of technical requirements has also been identified at this last phase of the project. Additionally, this document provides the state-of-the-art analysis of the base technology sectors related to PolicyCLOUD project and that could possibly be exploited, along with a list of baseline technological tools and solutions incorporated in the overall platform. This deliverable is a valuable input for the design of the overall architecture of the platform, as it will help to refine the design of the final version of the architecture to cope with the updated descriptions of the use cases scenarios.

This is the third and last of a series of documents providing the State of the art & requirements analysis for the PolicyCLOUD platform. As it has been published in M23, November 2021, this last version updates the current list of the user and technical requirements, taking into consideration that the use cases are more mature and their relevant scenarios have been better defined. At that point, the overall architecture will need to be further refined and extended to cover the more advanced scenarios that were not taken into account at an earlier phase. This final version of the deliverable covers all the remaining aspects and corrects potential erroneous decisions or unnecessary requirements that might have been identified earlier, so that it can drive the final version of the overall architecture of the project.

References

- [1] A Survey of Data Fusion in Smart City Applications, Lau at al. ELSEVIER INFORMATION FUSION 2019
- [2] Real-time pricing by data fusion on networks," Izumi at al, IEEE Transactions on Industrial Informatics, vol. 14, no. 3, 2018
- [3] Iterative channel estimation using lse and sparse message passing for mmwave mimo systems, Huang at al., IEEE Transactions on Signal Processing, vol. 67, no. 1, 2019.
- [4] Feature learning and analysis for cleanliness classification in restrooms, Jayasinghe at al, IEEE Access, vol. 7, 2019
- [5] Understanding the lifestyle of older population - Mobile crowdsensing approach, Marakkalage at al., IEEE Transactions on Computational Social Systems, vol. 6, no. 1, 2019
- [6] Harnessing multi-source data about public sentiments and activities for informed design, You at al., IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 2, 2018
- [7] Internet of things for green building management - Disruptive innovations through low-cost sensor technology and artificial intelligence, Tushar at al., IEEE Signal Processing Magazine, vol. 35, no. 5, 2018
- [8] A. Solanas et al., "Smart Health - A Context-Aware Health Paradigm within Smart Cities," IEEE Comm. Mag., vol. 52, no. 8, 2014, pp. 74–81, 2014.
- [9] Melnik S., Decker S., "A Layered Approach to Information Modeling and Interoperability on the Web," Proc. ECDL 2000 Workshop Semantic Web, <http://infolab.stanford.edu/~melnik/pub/sw00/sw00.pdf>, 2000.
- [10] New European Interoperability Framework, https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf , Accessed 2 Mar 2020.
- [11] Hellmann S., Lehmann J., Auer S., Brümmer M., Integrating NLP using Linked Data, 12th International Semantic Web Conference, pp. 21-25, 2013.
- [12] Fernandez, R.C., Mansour, E., Qahtan, A.A., et al., Seeping semantics - Linking datasets using word embeddings for data discovery, In - 2018 IEEE 34th International Conference on Data
- [13] Farias T. M., Stockinger K., Dessimo C, VoIDext - Vocabulary and Patterns for Enhancing Interoperable Datasets with Virtual Links, 2019.
- [14] Binding C., May K., Tudhope D., Semantic Interoperability in Archaeological Datasets - Data Mapping and Extraction Via the CIDOC CRM. In - Christensen-Dalsgaard B., Castelli D., Ammitzbøll Jurik B., Lippincott J. (eds) Research and Advanced Technology for Digital Libraries. ECDL 2008. Lecture Notes in Computer Science, vol 5173. Springer, Berlin, Heidelberg (2008)

- [15] Colpaert P., Van Compernelle M., et. al., Quantifying the Interoperability of Open Government Datasets, DOI - 10.1109/MC.2014.296, 2014.
- [16] JSON-LD - JSON for Linking Data, <http://json-ld.org> , Accessed 5 Mar 2020.
- [17] Xin, J., Afrasiabi, C., Lelong, S., Adesara, J., Tsueng, G., Su, A. I., Wu, C., Cross-linking BioThings APIs through JSON-LD to facilitate knowledge exploration, BMC bioinformatics, 19(1), 30. <https://doi.org/10.1186/s12859-018-2041-5> ,(2018)
- [18] International Organization for Standardization, “ISO/IEC/IEEE 29148:2011 – Systems and software engineering — Life cycle processes — Requirements engineering,” ISO/IEC/IEEE, Nov. 2011.
- [19] Wikipedia (17 March 2020). Comparison of JavaScript charting libraries, https://en.wikipedia.org/wiki/Comparison_of_JavaScript_charting_libraries, retrieved 2020-03-25)
- [20] Barot, Saurabh (2020). Top JavaScript Chart Libraries to Use in 2020 for Better Data Visualization, <https://aglowiditsolutions.com/blog/top-javascript-chart-libraries/>, retrieved 2020-03-25)
- [21] Borovikov, Ruslan (April 27th 2019). Top 10 JavaScript Charting Libraries for Every Data Visualization Need, <https://hackernoon.com/10-javascript-charting-libraries-data-visualization-b77523d23372>, retrieved 2020-03-26)
- [22] Puszynski, Arthur (15 May 2019). These are the best JavaScript chart libraries for 2019, <https://www.freecodecamp.org/news/these-are-the-best-javascript-chart-libraries-for-2019-29782f5e1dc2/>, retrieved 2020-03-26)
- [23] Shealy, Matt (21 Nov 2019). Best Data Visualization Tools for 2020 Reviewed, <https://readwrite.com/2019/11/21/best-data-visualization-tools-for-2020-reviewed/>, retrieved 2020-03-26)
- [24] Goel, Aman (29 Feb, 2020). 10 Best JavaScript Frameworks to Use in 2020, <https://hackr.io/blog/best-javascript-frameworks>, retrieved 2020-03-26)
- [25] Sviatoslav A. (JAN 08, 2020). The Best JS Frameworks for Front End, <https://rubygarage.org/blog/best-javascript-frameworks-for-front-end>, retrieved 2020-03-26)
- [26] Apache Drill – Schema-free SQL Query Engine for Hadoop, NoSQL and Cloud Storage, <https://drill.apache.org/>
- [27] Apache Impala, <http://impala.apache.org/>
- [28] M. Armbrust, R. Xin, C. Lian, Y. Huai, D. Liu, J. Bradley, X. Meng, T. Kaftan, M. Franklin, A. Ghodsi, M. Zaharia, “Spark SQL - relational data processing in Spark”, ACM SIGMOD, 2015, pp. 1383-1394.
- [29] F. Bugiotti, D. Bursztyn, A. Deutsch, I. Ileana, I. Manolescu, “Invisible glue - scalable self-tuning multi-stores”, in Conference on Innovative Data Systems Research (CIDR), 2015.

- [30] S. Dasgupta, K. Coakley, A. Gupta, “Analytics-driven data ingestion and derivation in the AWESOME polystore”, in IEEE International Conference on Big Data, 2016, pp. 2555-2564.
- [31] J. Duggan, A. J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, S. Zdonik, “The BigDAWG polystore system”, SIGMOD Record, vol. 44, no. 2, pp. 11-16, 2015.
- [32] V. Gadepally, P. Chen, J. Duggan, A. J. Elmore, B. Haynes, J. Kepner, S. Madden, T. Mattson, M. Stonebraker, “The BigDawg polystore system and architecture”, in IEEE High Performance Extreme Computing Conference (HPEC), 2016, pp. 1-6.
- [33] Y. Khan, A. Zimmermann, A. Jha, D. Rebholz-Schuhmann, R. Sahay, “Querying web pol-ystores”, in IEEE International Conference on Big Data, 2017.
- [34] Z. Minpeng, R. Tore, “Querying combined cloud-based and relational databases”, in Int. Conf. on Cloud and Service Computing (CSC), 2011, pp. 330-335.
- [35] K. W. Ong, Y. Papakonstantinou, and R. Vernoux, “The SQL++ semi-structured data model and query language - a capabilities survey of SQL-on-Hadoop, NoSQL and NewSQL databases”, CoRR, abs/1405.3631, 2014.
- [36] T. Özsu, P. Valduriez, Principles of Distributed Database Systems, 3rd ed. Springer, 2011, 850 pages.
- [37] Presto – Distributed Query Engine for Big Data, <https://prestodb.io/>
- [38] A. Simitsis, K. Wilkinson, M. Castellanos, U. Dayal, “Optimizing analytic data flows for multiple execution engines”, in ACM SIGMOD, 2012, pp. 829-840.
- [39] A. Tomasic, L. Raschid, P. Valduriez, “Scaling access to heterogeneous data sources with DISCO”, IEEE Trans. On Knowledge and Data Engineering, vol. 10, pp. 808-823, 1998.
- [40] J. Wang, T. Baker, M. Balazinska, D. Halperin, B. Haynes, B. Howe, D. Hutchison, S. Jain, R. Maas, P. Mehta, D. Moritz, B. Myers, J. Ortiz, D. Suciu, A. Whitaker, S. Xu, “The Myria big data management and analytics system and cloud service”, in Conference on Innovative Data Systems Research (CIDR), 2017.
- [41] Banks, G. (2009), Evidence-based policy making - What is it? How do we get it? (ANU Public Lecture Series, presented by ANZSOG, 4 February), Productivity Commission, Canberra.
- [42] Howlett, M. & Ramesh, M. (2003), Studying public policy - Policy cycles and policy subsystems. Toronto, ON - Oxford University Press Canada.
- [43] Big Policy Canvas. D5.2 Roadmap for Future Research Directions – Final Version. Big Policy Canvas, Francesco Mureddu. 2019.
- [44] Shroff ZC, Javadi D, Gilson L, Kang R, Ghaffar A. Institutional capacity to generate and use evidence in LMICs - current state and opportunities for HPSR. Health Res Policy Syst. 2017 Nov 9;15(1):94.

- [45] Zardo P, Collie A. Type, frequency and purpose of information used to inform public health policy and program decision-making. *BMC Public Health*. 2015 Apr 15;15:381.
- [46] O'Donoghue Jenkins L, Kelly PM, Cherbuin N, Anstey KJ. Evaluating and Using Observational Evidence - The Contrasting Views of Policy Makers and Epidemiologists. *Front Public Health*. 2016;4:267.
- [47] Lee BY, Ferguson MC, Hertenstein DL, Adam A, Zenkov E, Wang PI, et al. Simulating the Impact of Sugar-Sweetened Beverage Warning Labels in Three Cities. *Am J Prev Med*. 2018 Feb;54(2):197–204.
- [48] Powell-Wiley TM, Wong MS, Adu-Brimpong J, Brown ST, Hertenstein DL, Zenkov E, et al. Simulating the Impact of Crime on African-American Women's Physical Activity and Obesity. *Obes Silver Spring Md*. 2017 Dec;25(12):2149–55.
- [49] Atkinson J-A, Knowles D, Wiggers J, Livingston M, Room R, Prodan A, et al. Harnessing advances in computer simulation to inform policy and planning to reduce alcohol-related harms. *Int J Public Health*. 2018;63(4):537–46.
- [50] Freebairn L, Atkinson J-A, Kelly PM, McDonnell G, Rychetnik L. Decision makers' experience of participatory dynamic simulation modelling - methods for public health policy. *BMC Med Inform Decis Mak* [Internet]. Available from - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291959/>
- [51] Katrakazas P, et al. Public Health Policy Decision Models (PHPDM) v1, Deliverable D3.1 to the EVOTION-727521 Project funded by the European Union, ICCS
- [52] Kyriazis D, Autexier S, Brondino I, Boniface M, Donat L, Engen V, et al. CrowdHEALTH - Holistic Health Records and Big Data Analytics for Health Policy Making and Personalized Health. *Stud Health Technol Inform*. 2017;238:19–23
- [53] Moutselos K, Kyriazis D, Diamantopoulou V, Maglogiannis I. Trustworthy data processing for health analytics tasks. In - 2018 IEEE International Conference on Big Data. 2018. p. 3774–9.
- [54] Moutselos K, Maglogiannis I. Evidence-based Public Health Policy Models Development and Evaluation using Big Data Analytics and Web Technologies. *Medical Archives*. 2020; 74(1) - 47-53.
- [55] Kamateri E, Panopoulou E, Tambouris E, Tarabanis K, Ojo A, Lee D, et al. A Comparative Analysis of Tools and Technologies for Policy Making. In - Janssen M, Wimmer MA, Deljoo A, editors. *Policy Practice and Digital Science - Integrating Complex Systems, Social Simulation and Public Administration in Policy Research*. Cham - Springer International Publishing; 2015. p. 125–56.
- [56] P. J. Giabbanelli and M. Baniukiewicz, "Navigating Complex Systems for Policymaking Using Simple Software Tools," in *Advanced Data Analytics in Health*, Springer, Cham, 2018, pp. 21–40.