CLOUD FOR DATA-DRIVEN POLICY MANAGEMENT

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

D1.2 DATA MANAGEMENT PLAN

Dissemination Level	PU
Due Date of Deliverable	30/06/2020, Month 6
Actual Submission Date	29/06/2020
Work Package	WP1 Project Coordination
Task	T1.2
Type	ORDP: Open Research Data Pilot
Approval Status	
Version	V0.6
Number of Pages	p.1 – p.26

Abstract: This document provides the Data Management Plan for the PolicyCLOUD project according the Open Research Data Pilot guidelines and definition.
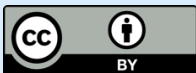
Versioning and Contribution History

Version	Date	Reason	Author
0.1	20/04/2020	Initial table of contents	Ricard Munné (ATOS)
0.2	27/05/2020	Initial content for sections Data Management Strategy and Project Datasets. Request for contributions and comments to partners.	Ricard Munné (ATOS)
0.3	15/06/2020	Complete version ready for Peer Review	Ricard Munné (ATOS)
0.4	17/06/2020	Address peer review comments and suggestions	Armend Duzha (MAG), Ricard Munné (ATOS)
0.5	22/06/2020	Address peer review comments and suggestions	Enol Fernández (EGI), Ricard Munné (ATOS)
0.6	29/06/2020	Quality Check. Document ready for submission.	Argyro Mavrogiorgou (UPRC)

Author List

Organisation	Name
ATOS	Ricard Munné

Abbreviations and Acronyms

Abbreviation/Acronym	Definition
ATOS	Atos Spain SA
CA	Consortium Agreement
CAP	Common Agricultural Policy
CMS	Content Management System
CO	Confidential
CrowdHEALTH	Collective Wisdom Driving Public Health Policies (H2020 project)
CSV	Comma Separated Values
DataCite	International non-for-profit organisation for data citation
DMP	Data Management Plan
Drupal	Free and open source CMS
Dublin Core	Dublin Core Metadata Initiative
Dx.y	Deliverable number y belonging to Activity x
DWF	DWF Germany Rechtsanwaltsgesellschaft MBH
EC	European Commission
EGI	Stichting EGI
EO4AGRI	Earth Observation for Agricultural (H2020 project)
EU	European Union
FAIR	Findable, accessible, interoperable and re-usable
GA	Grant Agreement
IBM	IBM Israel – Science and Technology LTD
GA	Grant Agreement
GDPR	General Data Protection Regulation
H2020	Horizon 2020
Hec	Hectare
ICT	Information and Communication Technology
KPI	Key Performance Indicator
ICCS/NTUA	Institute of Communications and Computer Systems
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LXS	Leanxcale SL
M	Month
MAG	Maggioli
MARCXML	Machine Readable Cataloging Extensible Markup Language
ODF	Open Document Format for Office Applications
OpenAIRE	European project supporting Open Science with a technical infrastructure harvesting research output from connected data providers.
ORDP	Open Research Data Pilot
ownCloud	ownCloud Enterprise File Sync and Share Software
PDF	Portable Document Format
PolicyCLOUD	Policy Management through technologies across the complete data lifecycle on cloud environments
R&I	Research and Innovation
SARGA	Sociedad Aragonesa de Gestión Agroambiental
SIGPAC	Sistema de Información Geográfica de parcelas agrícolas (Geographic Information System for agricultural parcels)
SME	Small and Medium Enterprises
SOF	Stolichna Obshtina
SQL DB	Structured Query Language Data Base
Tx.y	Task number y belonging to WP x
TRUST-IT	Trust-It Services Limited

Abbreviation/Acronym	Definition
UBI	Ubitech Limited
UC	Use Case
XML	Extensible Markup Language
WP	Work Package
Zenodo	An all-purpose open research repository created by OpenAIRE and CERN

Contents

List of Tables

List of Figures

Executive Summary

This deliverable is the initial version of the Data Management Plan (DMP) for the PolicyCLOUD project and it provides a first description of the datasets available according to FAIR principles. The overall purpose of this document is to support the data management lifecycle for all data that will be collected, processed or generated by the project in order to maximise its access, according to the H2020 Pilot on Open Research Data(ORDP) in which the project participates.

A Data Management strategy is defined to provide the general rules and mechanisms for the access management of project data. Each dataset in the project is identified and described and information is provided on to which extent it is standard compliant, and how the data will be available, accessible interoperable and reusable.

A total of 17 datasets have been identified in this first version of the deliverable, being most of them data that will be collected and processed by the four pilot use cases. Other datasets are mostly stakeholders' information and public deliverables produced within the project scope.

This DMP is a living document and will be updated during the course of the project in new versions at M12, M24 and M36.

1 Introduction

1.1 Purpose of the document

The purpose of the PolicyCLOUD Data management Plan (DMP) is to support the data management life cycle for all data that will be collected, processed or generated by the project according to the Guidelines on FAIR Data Management in Horizon 2020 [1] .

PolicyCLOUD Data Management Plan takes as starting point the Data Management Plan from EO4AGRI [2] and CrowdHEALTH [3] projects (where ATOS is also involved) and adapts them for this project specific needs.

1.2 Relation to other project work

This DMP has impact on the actions that will be performed over datasets that will be collected, processed or generated. Specific impact is expected over the data collected and generated in WP1 Project Coordination, over data managed by the pilots in WP6, including data used or generated for the policies management and also from the co-creation workshops at the pilot sites, and finally over Task 7.3 for the Communication and Stakeholder engagement .

1.3 Structure of the document

This document is structured in three major chapters:

Section 2 defining the data management strategy that will be followed by the project according to the Guidelines on FAIR Data Management in Horizon 2020 [1].

Section 3 presents the details about the datasets managed by the project including the main information to handle them regarding data sharing, archiving and security.

Section 4 presents the main conclusions and next steps regarding this Data Management Plan.

1.4 Identification of datasets

The datasets have been initially identified through the requirements collection, reported in D2.1 State of the art & Requirements Analysis, which are input datasets for the pilots. Additionally, datasets managed by other Tasks in the project as described in section 1.2 above, for the Project coordination and Communication work packages.

Additional datasets will be included in the next steps of the project in next versions of this DMP, as data collected and outputs from the pilots are clearly defined, which may be in turn be published through the Data Marketplace developed in WP7.

2 Data Management strategy

PolicyCLOUD project involves the processing of data from different sources to provide evidence-based policy making in the full lifecycle of policy management. In this section it is described the Data Management strategy of PolicyCLOUD, based on the guidelines on FAIR Data Management in Horizon 2020 [1] and on the template for FAIR Data Management in H2020 [4], with regards to the Findable, Accessible, Interoperable and Re-usable principles.

This strategy is set according to the following principles:

- Identification of datasets collected or generated by the project.
- Definition of the principles for exploitation, availability, access rights and re-use of data managed by the project.
- Definition of the principles for data archiving and preservation.
- Definition of the principles for ethical and legal compliance.

2.1 Type of data to be generated/collected/extracted

2.1.1 Data Source and Acquisition

PolicyCLOUD project mainly deals with the data collection from different sources that will feed the four pilots use cases. In addition, other sources of data are the co-creation workshops at the pilot sites, and the PolicyCLOUD communication and dissemination activities.

- **Collection of data for pilots**: Pilots will be demonstrated and validated through specific use cases for Evidence Based Policy Making. The uses cases will define a set of policies and KPIs based on specific input data. This data will be collected from different sources, including social networks, Open Data, Public Datasets, data collected by the public Organisations running the use cases, and other sources not yet classified. In this regard, Task 6.2 Use cases Definition & Design, deals with the definition of the use cases and their interfaces, while Task 2.1 deals with the identification of data to be collected. Task 3.5 Ethical and legal compliance framework provides analysis of legal, regulatory, societal and ethical aspects to ensure that these aspects are considered during the design and development phases, and specially to everything surrounding data management in the pilots.
- **Co-creation workshops at the pilot sites**: Under Task 6.2 Use Cases Definition and Design and Task 6.3 Use Cases Implementation & Experimentation, a series of co-creation workshops will be organised involving end user organisations (i.e. public authorities and organisations) and other relevant stakeholders related to the use cases. These workshops will provide guidance, valuable insights/advices and feedback during the development, implementation and testing of the PolicyCLOUD platform and services in each use case. The results from these events will be collected and documented in small reports that will be made available to the technical partners and the research community.
- **PolicyCLOUD communication and dissemination activities**: This includes data generated and collected in WP1 Project Coordination, and in Task 7.3 Communication & Stakeholder Engagement. In WP1 this is related to the user accounts and profile data required for internal administrative tasks like managing the project's mailing lists and the document repository in ownCloud[1]. Task 7.3 involves the

---

[1] https://owncloud.org/

organisation of joint events and workshops and building a stakeholders' database of profiled contacts. This requires the management of external users' personal data to manage event invitations, newsletters distribution, etc, where social media networks may be used for communication (including LinkedIn, Twitter and YouTube). Information will be captured at the PolicyCLOUD website (https://policycloud.eu/) through the users' self-registration form. This data will be stored in a private repository under data protection measures and according to the Privacy Policy available at the PolicyCLOUD website.

2.1.2 Type of data

Data generated or collected by the project can be classified as:

- No personal data: that corresponds to data not related to any individual, either project collaborator, external stakeholder or any other identifiable person, and therefore not affected by Data Protection regulations. Examples of this are the Wine register or the Production data per grape variety datasets, to be used in UC#2.
- Personal data: is data that can be related to any individual that may be identified
  - Directly from the data
  - Indirectly through the data and additional information that is available in the environment that can provide re-identification of that individual.

In this regard, PolicyCLOUD will collect personal data only when it is strictly necessary, and primarily for the management of activities in the project. Regarding the data used in the use cases by the pilots, and stored in the PolicyCLOUD database, the PolicyCLOUD architecture is incorporating a series of mechanisms to guarantee legal compliance and to keep privacy and respect licensing terms of the original data sources.

An ownCloud server is used as project repository for documents exchanged within the consortium. It provides the support needed by the documentation storage, review process, information sharing, and work in groups by all partners in order to achieve the common goals of the project. All relevant information for the project is to be stored in this repository, including contractual documents (GA, CA), amendments, review-related documentation, reporting documentation, contact details, templates, working documents of deliverables, internal working documents, agendas, minutes, etc. Moreover, final versions of all deliverables are to be uploaded there, as detailed in the PolicyCLOUD Project Management Plan [5].

For further details on how the activities focused on the ethical and privacy issues and compliance to legislation are managed in PolicyCLOUD project, please refer to section 2.1.6.

2.1.3 Standards and metadata to be used

All PolicyCLOUD documents and data that will be made publicly accessible and will be published in the Zenodo PolicyCLOUD community[2]. In this regard, the Zenodo metadata will be used, which is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichments [6]. This metadata uses a JSON Schema[3] as an internal representation and provides the possibility to export to other formats like Dublin Core or MARCXML.

---

[2] https://zenodo.org/communities/policycloud
[3] https://zenodo.org/schemas/records/record-v1.0.0.json

2.1.4 Exploitation, availability of data, access, sharing, and re-use

As many PolicyCLOUD results as possible will be made openly accessible according to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 [9]. These will be made available through an Open Access repository, and those produced as direct outcomes of the project will be published as detailed in section 2.1.3 above.

Open access to peer-reviewed scientific publications will be performed by publishing either in green or gold open access journals, and announced on the PolicyCLOUD website, the OpenAIRE portal[4] and in the R&I Participants Portal[5]. The process can be automated by depositing the publication in an OpenAIRE compliant repository by the authors in case of green open access or by the scientific publisher in the case of gold open access.

In case the repository is not fully OpenAIRE compliant, the publication or data must be linked at https://www.openaire.eu/participate/claim with the European Commission as funding agency.

For the dissemination of PolicyCLOUD results, green open access without publication fees are preferred. Journal publications must guarantee permission to self-archive the published article at most 6 months after publication. However, due to the journal visibility and acceptance it may be considered to publish in a gold open access, made available as open access by the publisher. In this case, publication fees are eligible and can be reimbursed as project costs by the PolicyCLOUD partners.

As for the re-use and interoperability of the data, reports and open data produced in the course of the project will be made available in commonly used formats, as ODF or PDF for documents and JSON or XML for data. In addition, the following standard vocabularies will be used in the default metadata schema for all types of open data:

- License: Open Definition[6]
- Funders: FundRef[7]
- Grants: OpenAIRE[8]

2.1.5 Archiving and preservation

For the project documentation, including deliverables intermediate versioning, meetings documentation, including agendas, minutes and presentations, and for any other document used for the development of the project, an ownCloud server is hosted by ATOS to provide a collaborative environment for content sharing across consortium partners, and therefore reducing the need for email sharing of documents. The server is only accessible for consortium team members who have registered and signed the Data Protection Form, as described in the PolicyCLOUD Project Management Plan [5].

For data produced as a result of the use cases demonstration, this data will be stored in the PolicyCLOUD data store and made available as described in section 2.1.4 above.

---

[4] https://www.OpenAIRE.eu/
[5] https://ec.europa.eu/research/participants
[6] http://opendefinition.org/
[7] https://www.crossref.org/services/funder-registry
[8] http://api.OpenAIRE.eu/

2.1.6 Ethics and legal compliance

Ethical and legal issues are considered a transversal challenge in PolicyCLOUD. For this reason, as described in PolicyCLOUD Grant Agreement [7], an ethical framework is considered as one of the architecture building blocks, that includes legal, security and ethical aspects, to ensure compliance to the corresponding requirement, while maximizing the societal acceptability and trust in PolicyCLOUD solutions. More specifically, in WP3 for *Cloud Infrastructures Utilization & Data Governance*, there is a specific Task, T3.5 for the implementation of such *Ethical and legal compliance framework*. This work will be reported under *PolicyCLOUD's Societal and Ethical Requirements & Guidelines*, in three issues, corresponding to deliverables D3.3 in M10, D3.6 in M22 and D3.9 in M34.

This framework will seek an optimal fit of these aspects into the PolicyCLOUD platform design. More specifically on the ethical and social impacts derived from the use of social networks data, which is already considered in at least two of the four piloting use cases, and for which the input datasets are described in section 3 of this report. This of course takes into consideration the management of personal data and the solutions to minimise the need for such data in the data to be analysed, not only for project results to be shared but also to ensure compliance with data protection regulations of the platform being developed, and more specifically for the Data Marketplace which is another of the architecture building blocks.

The ethical and legal principles that will guide PolicyCLOUD project are declared in section 5 "Ethics and Security" of the PolicyCLOUD GA [7], including privacy, for the protection of personal data, and security enforcement within the project.

In deliverable D8.1 [8], corresponding to the Ethics requirements, it is reported the approach for informed consent procedures in PolicyCLOUD, where it is declared that PolicyCLOUD is not aiming at processing personal data of external participants in the course of the project.

2.1.7 Resources

No costs are foreseen to make the resulting data and reports generated by PolicyCLOUD openly available. It is planned to use Zenodo repository for the publication. In Figure 1, the link to PolicyCLOUD Zenodo page at the PolicyCLOUD website can be appreciated. In Figure 2 a sample of the Zenodo PolicyCLOUD community[9] is shown.
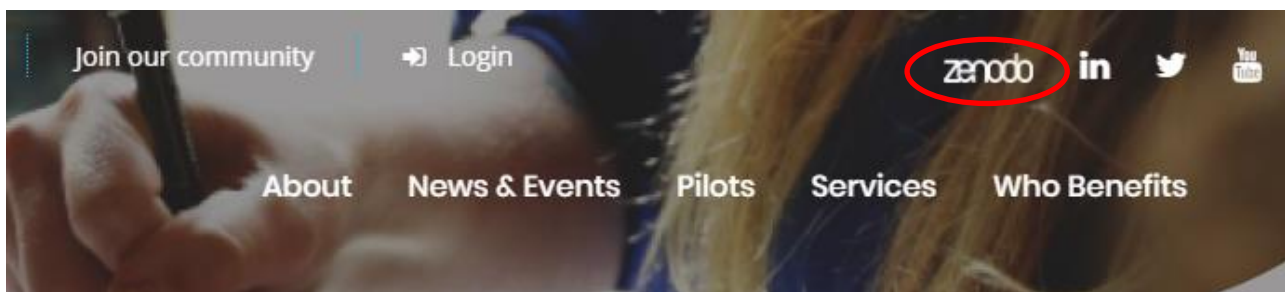
FIGURE 1 - ZENODO LINK AT POLICYCLOUD SITE

---

[9] https://zenodo.org/communities/policycloud
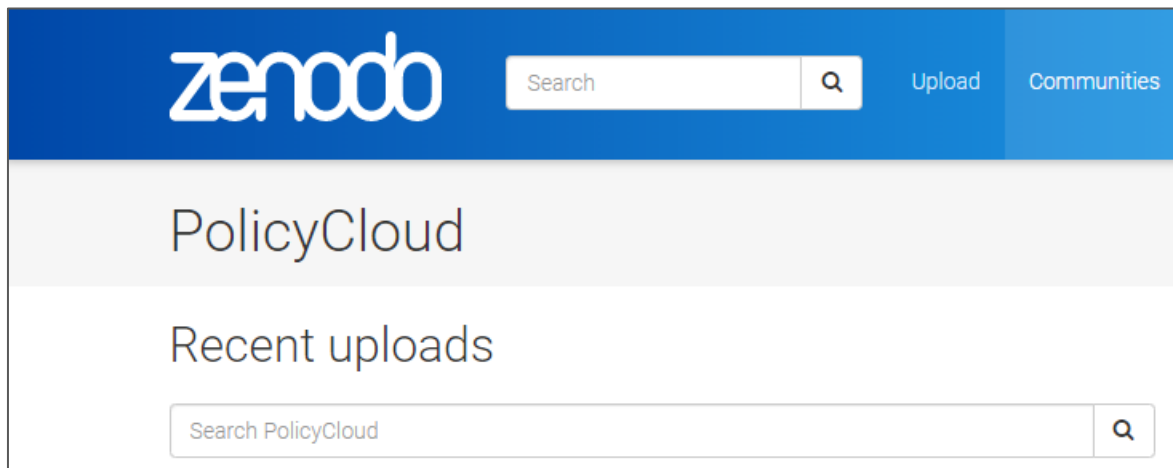
FIGURE 2 - POLICYCLOUD COMMUNITY PAGE AT ZENODO

DATA MANAGEMENT IN THE CONSORTIUM

Data management activities are responsibility of the whole consortium, from creation of data, publication and dissemination, following the Data Management Plan, Project Management Plan [5] and Communication and Dissemination Strategy (to be delivered in M12) guidelines.

The **Project Coordinator** is responsible for:

- Develop the data management plan in collaboration with the consortium partners.
- Supervise of the data management activities (for collection and publication) and associated milestones in coordination with Workpackage managers.
- Write the data management plan (D1.2) releases (in M6, M12, M24 and M36).
- Provide support to arising issues.

The **Work Package Leaders** are responsible for:

- Implementing the Data Management guidelines in their respective WPs.
- Monitor the Data Management activities and milestones with partners.
- Provide input to the data management plan by analysing and summarising the WP-specific related data.
- Provide support and guidance for the publication of open data and documents.
- Monitor that open results (data and software) are deposited in the default repository or a complementary OpenAIRE-compliant repository and sending reminders to partners.
- Monitor that open results available in OpenAIRE are properly linked with PolicyCLOUD (see https://www.OpenAIRE.eu/participate/claim).
- Coordinate with the Project Management for any ethical or privacy issues that may prevent the publication of data.

The **Communication task leader** is responsible for:

- Provide support on the best publication path (green or gold open access).

- Provide support for publication on scientific publications.

- Ensure that the open access policy of the journal complies with the H2020 open data requirements [9] before the paper is submitted.

- Monitor that green access (self-archiving) publications are deposited in repositories.

- Monitor that metadata about publications is made available in the R&I Participant Portal (preferably automatically through OpenAIRE) and on the PolicyCLOUD website.

- Ensure that research data related to a publication is made available in repositories and linked to respective publication.

- Monitor possible embargo periods and remind to partners.

- Ensure that publications available in OpenAIRE are properly linked with PolicyCLOUD (see https://www.OpenAIRE.eu/participate/claim).

The **Data Provider / Scientist** is responsible for:

- Informing data & dissemination leaders when new open data / papers ready for publication are available.

- Describe the data (through appropriate metadata) or scientific publication in accordance to the PolicyCLOUD data management policy (e.g. according to the chosen metadata standard).

- Depositing (publishing into a repository) the data or scientific publication in accordance to the PolicyCLOUD data management policy and with help of the tools (catalogue, repository, etc.) provided by the project.

3 Project Datasets

3.1 Dataset 1: PolicyCLOUD Stakeholder registered in PolicyCLOUD website

Dataset 1: PolicyCLOUD Stakeholder Database from project partners	
Information	**Description**
Dataset name	PolicyCLOUD Stakeholder Database registered in website.
Dataset description	Profiled Stakeholder database collected from stakeholders self-registered in the PolicyCLOUD project website to contact and provide project information and invite for project activities.
Standards and metadata	Website database from Drupal CMS.
Data sharing	Data will not be openly shared. Shared only with consortium partners. The dataset will not be made open as its sole purpose is to keep track of relevant contacts for PolicyCLOUD stakeholders to communicate, disseminate and collect feedback for project results.
Archiving and preservation	Website database from Drupal CMS.
Security needs /Legal issues	The dataset is private and used only for internal purposes of the project, therefore its access remains restricted solely for the project partners.

TABLE 1 - DATASET POLICYCLOUD STAKEHOLDER DATABASE REGISTERED IN WEBSITE

3.2 Dataset 2: PolicyCLOUD Stakeholder Database from project partners

Dataset 1: PolicyCLOUD Stakeholder Database from project partners	
Information	**Description**
Dataset name	PolicyCLOUD Stakeholder Database from project partners.
Dataset description	Profiled Stakeholder database with contact from partners networks and desktop research to contact and invite for project activities. This profiled stakeholder database holds all relevant project partner connections to: • Policy makers & public administrations • Industry (large enterprise and SME, data solution providers, cloud solution providers….) • R&I (projects engaged in) • Citizens (local interest groups etc.) No direct stakeholder contact data is stored, only reference to the project partner who has direct access to the stakeholder.
Standards and metadata	Excel document
Data sharing	Data will not be openly shared. Shared only with consortium partners. The dataset will not be made open as its sole purpose is to keep track of relevant contacts for PolicyCLOUD stakeholders to communicate, disseminate and collect feedback for project results.
Archiving and preservation	Data is stored in the ownCloud PolicyCLOUD repository.

Dataset 1: PolicyCLOUD Stakeholder Database from project partners	
Information	**Description**
Security needs /Legal issues	The dataset is private and used only for internal purposes of the project, therefore its access remains restricted solely for the project partners.

TABLE 2 - DATASET POLICYCLOUD STAKEHOLDER DATABASE FROM PROJECT PARTNERS

3.3 Dataset 3: PolicyCLOUD Public Deliverables

Dataset 1: PolicyCLOUD Public Deliverables	
Information	**Description**
Dataset name	PolicyCLOUD Public Deliverables.
Dataset description	PolicyCLOUD is an Innovation Action that provides an integrated environment for evidence-based policy making based on a cloud environment. So, it will develop an innovative approach, and several reports will be produced documenting the project achievements. For a detailed list of public deliverables, please check Annex I – PolicyCLOUD public project deliverables.
Standards and metadata	Microsoft Word document and PDF format. Zenodo deposition metadata domain model.
Data sharing	Data to be openly shared
Archiving and preservation	The dataset is stored in PolicyCLOUD private ownCloud and public Zenodo repository.
Security needs /Legal issues	N/A.

TABLE 3 - DATASET POLICYCLOUD PUBLIC DELIVERABLES

3.4 Dataset 4: Participatory policies against radicalization in Twitter (UC#1)

Dataset 1: Participatory policies against radicalization in Twitter	
Information	**Description**
Dataset name	Twitter posts for radicalisation use case.
Dataset description	Relevant Twitter posts published by users, captured and processed for subsequent analysis in UC#1.
Standards and metadata	Text and images in JSON format.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 4 - DATASET TWITTER POSTS FOR RADICALISATION USE CASE

3.5 Dataset 5: Relevant posts published by users in Reddit (UC#1)

Dataset 2: Participatory policies against radicalization in Reddit	
Information	**Description**
Dataset name	Reddit posts for radicalisation use case.
Dataset description	Relevant Reddit posts published by users, captured and processed for subsequent analysis in UC#1.
Standards and metadata	Text and images in JSON format.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 5 - DATASET REDDIT POSTS FOR RADICALISATION USE CASE

3.6 Dataset 6: Relevant news available on the web (UC#1)

Dataset 3: Participatory policies against radicalization from news posted in websites	
Information	**Description**
Dataset name	News posted in websites for radicalisation use case.
Dataset description	Relevant news published in websites, captured and processed for subsequent analysis in UC#1.
Standards and metadata	Unstructured information, text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 6 - DATASET NEWS POSTED IN WEBSITES FOR RADICALISATION USE CASE

3.7 Dataset 7: Global Terrorism Database (UC#1)

Dataset 4: Global Terrorism Database	
Information	**Description**
Dataset name	Global Terrorism Database.
Dataset description	Open-source database including information on domestic and international terrorist attacks around the world from 1970 through 2018, and now includes more than 190,000 cases. (https://www.start.umd.edu/gtd/)
Standards and metadata	Structured information, text.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.

Dataset 4: Global Terrorism Database	
Information	**Description**
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 7 - DATASET GLOBAL TERRORISM DATABASE

3.8 Dataset 8: Common Agricultural Policy (UC#2)

Dataset 5: Common Agricultural Policy	
Information	**Description**
Dataset name	Common Agricultural Policy.
Dataset 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.
Standards and metadata	Semi-structured. Virtuoso (triplets), JSON or XML. https://opendata.aragon.es/datos/catalogo?texto=pac
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 8 - DATASET COMMON AGRICULTURAL POLICY

3.9 Dataset 9: Wine register (UC#2)

Dataset 6: Wine register	
Information	**Description**
Dataset name	Wine register.
Dataset description	SIGPAC reference, variety, cultivation year, area (Hec) form Aragon Government.
Standards and metadata	Structured. CSV. https://www.aragon.es/en/-/consultas-sigpac
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 9 - DATASET WINE REGISTER

3.10 Dataset 10: Production data per grape variety (UC#2)

Dataset 7: Production data per grape variety	
Information	**Description**
Dataset name	Production data.
Dataset description	Production data per grape variety.
Standards and metadata	Structured, SQL DB.
Data sharing	To be further analysed if source data will be shared.
Archiving and preservation	Data can be imported and stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 10 - DATASET PRODUCTION DATA

3.11 Dataset 11: Wine varieties and brands information from Twitter (UC#2)

Dataset 8: Wine varieties and brands information from Twitter	
Information	**Description**
Dataset name	Twitter posts for wine use case.
Dataset description	Relevant Twitter posts published by users about wine varieties, brands, captured and processed for subsequent analysis in UC#2.
Standards and metadata	Text and images in JSON format.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 11 - DATASET WINE VARIETIES AND BRANDS INFORMATION FROM TWITTER

3.12 Dataset 12: Wine varieties and brands information from Facebook (UC#2)

Dataset 9: Wine varieties and brands information from Facebook	
Information	**Description**
Dataset name	Facebook posts for wine use case.
Dataset description	Relevant Facebook posts published by users about wine varieties, brands, captured and processed for subsequent analysis in UC#2.
Standards and metadata	Unstructured information, text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.

Dataset 9: Wine varieties and brands information from Facebook	
Information	**Description**
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 12 - DATASET WINE VARIETIES AND BRANDS INFORMATION FROM FACEBOOK

3.13 Dataset 13: Wine varieties and brands information from Instagram (UC#2)

Dataset 10: Wine varieties and brands information from Instagram	
Information	**Description**
Dataset name	Instagram posts for wine use case.
Dataset description	Relevant Instagram posts published by users about wine varieties, brands, captured and processed for subsequent analysis in UC#2.
Standards and metadata	Unstructured information, text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 13 - DATASET WINE VARIETIES AND BRANDS INFORMATION FROM INSTAGRAM

3.14 Dataset 14: Wine varieties and brands information from LinkedIn (UC#2)

Dataset 11: Wine varieties and brands information from LinkedIn	
Information	**Description**
Dataset name	LinkedIn posts for wine use case.
Dataset description	Relevant LinkedIn posts published by users about wine varieties, brands, captured and processed for subsequent analysis in UC#2.
Standards and metadata	Text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 14 - DATASET WINE VARIETIES AND BRANDS INFORMATION FROM LINKEDIN

3.15 Dataset 15: Wine varieties and brands information from the web (UC#2)

Dataset 12: Wine varieties and brands information from websites	
Information	**Description**
Dataset name	Website news for wine use case.
Dataset description	Relevant web news published by users about wine varieties, brands, captured and processed for subsequent analysis in UC#2.
Standards and metadata	Unstructured information, text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 15 - DATASET WINE VARIETIES AND BRANDS INFORMATION FROM WEBSITES

3.16 Dataset 16: Sofia Municipality Signals (UC#3)

Dataset 13: Sofia Municipality Signals	
Information	**Description**
Dataset name	Sofia Municipality Signals.
Dataset description	Signals from citizens, coming through the contact centre of the municipality.
Standards and metadata	Structured information, text and images.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 16 - DATASET SOFIA MUNICIPALITY SIGNALS

3.17 Dataset 17: Unemployment Claimant Count (UC#4)

Dataset 14: Sofia Municipality Signals	
Information	**Description**
Dataset name	Unemployment Claimant Count.
Dataset description	Claimant Count unemployment measure, combining Jobseekers Allowance (JSA) claimants and Universal Credit claimants who are actively seeking work.
Standards and metadata	Structured information, CSV.
Data sharing	Source data will not be shared. To be further analysed if processed data will be shared.
Archiving and preservation	Data obtained from the processing of source data will be stored in the PolicyCLOUD Data Store.

Dataset 14: Sofia Municipality Signals	
Information	**Description**
Security needs /Legal issues	Source data will be processed according licensing terms and conditions from the source owner.

TABLE 17 - DATASET UNEMPLOYMENT CLAIMANT COUNT

4 Conclusions

This initial Data Management Plan (DMP) for the PolicyCLOUD project, reports the provisions for the data management strategy to be applied to the project datasets in general to make project data compliant with the FAIR guidelines.

PolicyCLOUD DMP procedures make use of solutions and standards like DataCite metadata, the OpenAIRE initiative and the Zenodo research data repository for their implementation. This is to ensure that PolicyCLOUD results, including public reports, open data, open access publications and open source software are accessible and available once the project is finished.

This DMP furthermore provides the first overview on the data that is collected, processed or generated following the methodology and standards set out in the data management strategy. The DMP is a living document and will be updated regularly with new data resulting from pilots' development and implementation, when more details about the data management for the known datasets are available or when the data management strategy needs to be updated.

References

[1] European Commission. (2016). *Guidelines on FAIR Data Management in Horizon 2020*. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Retrieved 2020-04-27.

[2] EO4AGRI. *D1.5 – Data Management Plan*. Miguel Ángel Esbrí. 2019.

[3] CrowdHEALTH. *D1.2 – Data Management Plan*. Rosana Valle. 2017.

[4] European Commission, (2016). *Template Horizon 2020 Data Management Plan (DMP)*, https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx. Retrieved 2020-04-27.

[5] PolicyCLOUD. *D1.1 - Project Management Plan*. Maria Teresa García. 2020.

[6] Zenodo. Principles. https://about.zenodo.org/principles/. Retrieved 2020-05-21.

[7] PolicyCLOUD. *PolicyCLOUD Grant Agreement number 879675*. European Commission. 2019.

[8] PolicyCLOUD. *D8.1 – POPD-Requirement No. 1*. Marc Meerkamp. 2020.

[9] European Commission. (2017). Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Retrieved 2020-05-21.

Annexes

Annex I – PolicyCLOUD public project deliverables

The following table summarizes all public deliverables produced by PolicyCLOUD project, indicating its number, name, work package, leading partner, type, dissemination level and submission plan to the EC, with M1 being January 2020. Many deliverables have versions submitted during project life as indicated.

Del. no.	Deliverable name	WP no.	Lead beneficiary	Type	Dissemin. level	Delivery date (Mx)
D1.2	Data Management Plan	WP1	ATOS	ORDP	Public	6, 12, 24, 36
D2.1	State of the Art & Requirements Analysis	WP2	LXS	Report	Public	6, 12, 22
D2.2	Conceptual Model & Reference Architecture	WP2	ICCS/NTUA	Report	Public	8, 18, 30
D3.1	Cloud Infrastructure Incentives Management and Data Governance: Design and Open Specification 1	WP3	UBI	Report	Public	8
D3.2	Cloud Infrastructure Incentives Management and Data Governance: Software Prototype 1	WP3	UBI	Demonstrator	Public	10
D3.3	PolicyCLOUD's Societal and Ethical Requirements & Guidelines - M10	WP3	DWF	Report	Public	10
D3.4	Cloud Infrastructure Incentives Management and Data Governance: Design and Open Specification 2	WP3	UBI	Report	Public	20
D3.5	Cloud Infrastructure Incentives Management and Data Governance: Software Prototype 2	WP3	UBI	Demonstrator	Public	22
D3.6	PolicyCLOUD's Societal and Ethical Requirements & Guidelines - M22	WP3	DWF	Report	Public	22
D3.7	Cloud Infrastructure Incentives Management and Data Governance: Design and Open Specification 3	WP3	UBI	Report	Public	32
D3.8	Cloud Infrastructure Incentives Management and Data Governance: Software Prototype 3	WP3	UBI	Demonstrator	Public	34
D3.9	PolicyCLOUD's Societal and Ethical Requirements & Guidelines	WP3	DWF	Report	Public	34
D4.1	Reusable Model & Analytical Tools: Design and Open Specification 1	WP4	IBM	Report	Public	8
D4.2	Reusable Model & Analytical Tools: Software Prototype 1	WP4	IBM	Demonstrator	Public	10
D4.3	Reusable Model & Analytical Tools: Design and Open Specification 2	WP4	IBM	Report	Public	20
D4.4	Reusable Model & Analytical Tools: Software Prototype 2	WP4	IBM	Demonstrator	Public	22
D4.5	Reusable Model & Analytical Tools: Design and Open Specification 3	WP4	IBM	Report	Public	32
D4.6	Reusable Model & Analytical Tools: Software Prototype 3	WP4	IBM	Demonstrator	Public	34

Policy Cloud
Cloud for Data-Driven Policy Management

Del. no.	Deliverable name	WP no.	Lead beneficiary	Type	Dissemin. level	Delivery date (Mx)
D5.1	Methods for using Cloud by Public Authorities	WP5	SOF	Report	Public	20, 32
D5.2	Cross-sector Policy Lifecycle Management: Design and Open Specification 1	WP5	MAG	Report	Public	8
D5.3	Cross-sector Policy Lifecycle Management: Software Prototype 1	WP5	MAG	Demonstrator	Public	10
D5.4	Cross-sector Policy Lifecycle Management: Design and Open Specification 2	WP5	MAG	Report	Public	20
D5.5	Cross-sector Policy Lifecycle Management: Software Prototype 2	WP5	MAG	Demonstrator	Public	22
D5.6	Cross-sector Policy Lifecycle Management: Design and Open Specification 3	WP5	MAG	Report	Public	32
D5.7	Cross-sector Policy Lifecycle Management: Software Prototype 3	WP5	MAG	Demonstrator	Public	34
D6.2	Integration of Results: PolicyCLOUD Complete Environment	WP6	ICCS/NTUA	Demonstrator	Public	12, 24, 36
D6.3	Use Case Scenarios Definition & Design	WP6	SARGA	Report	Public	8, 16, 28
D6.4	Use Cases Implementation & Experimentation	WP6	MAG	Demonstrator	Public	12, 24, 36
D6.5	Use Cases Adaptation & Recommendations	WP6	ITA	Report	Public	14, 24, 36
D7.1	Initial Publication Package	WP7	TRUST-IT	Report	Public	2
D7.4	Data Marketplace: Design and Open Specification	WP7	ATOS	Report	Public	20, 32
D7.5	Data Marketplace: Software Prototype	WP7	ATOS	Demonstrator	Public	22, 34
D7.6	Communication and Dissemination Strategy	WP7	TRUST-IT	Report	Public	12, 24, 36
D7.7	Standardisation Plan & Activities	WP7	TRUST-IT	Report	Public	24, 36