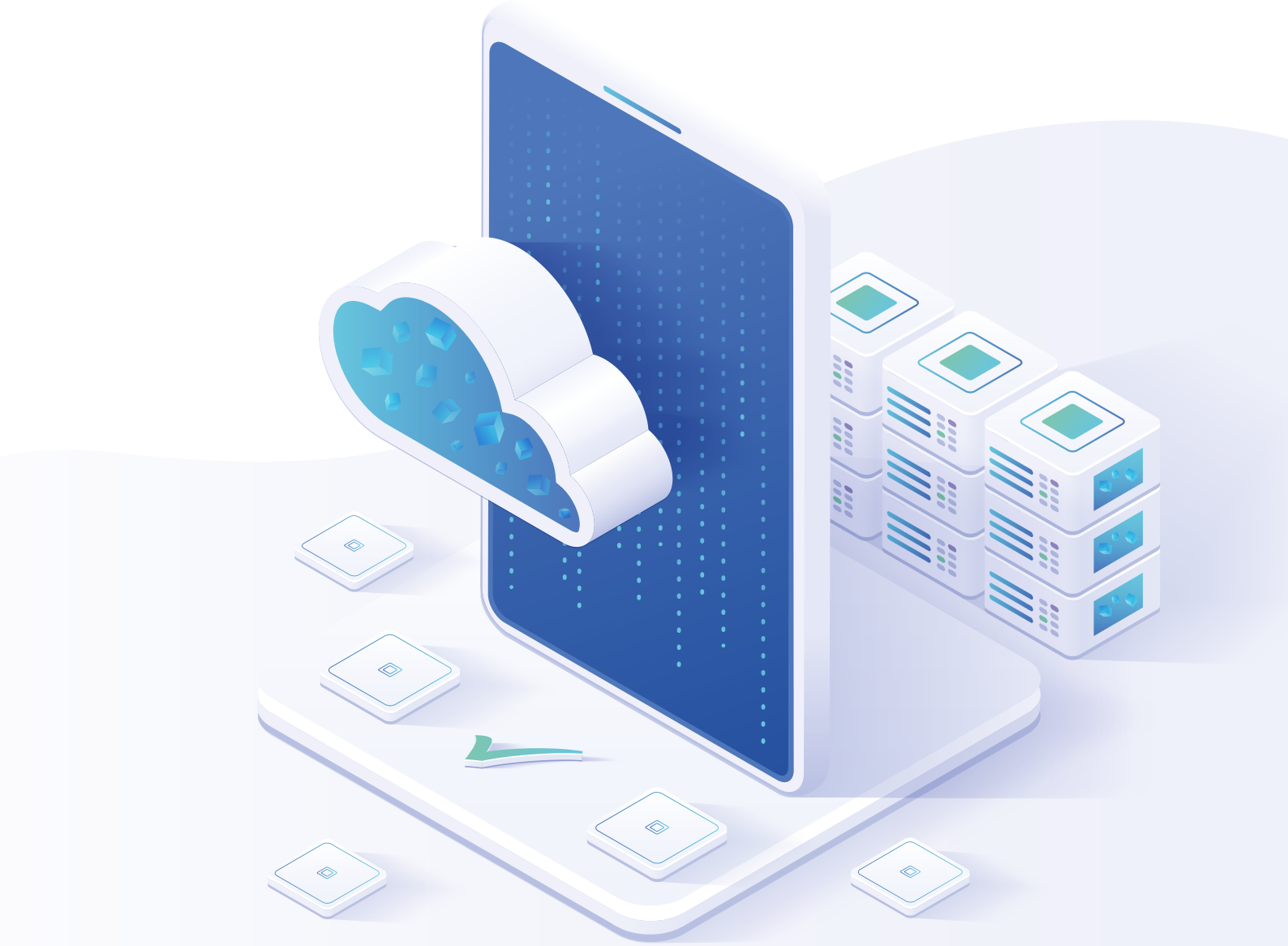
Policy Cloud
Cloud for Data-Driven Policy Management

In coordination with
StandICT.eu 2023
ICT STANDARDISATION OBSERVATORY AND SUPPORT FACILITY IN EUROPE

Cloud for **Data Driven Policy Management**

Research and Innovation Action

Towards interoperable Data Driven Policymaking



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

StandICT.eu 2023 Project
Project Number: 951972
Start Date of project: 01/09/2020
Duration: 36 months

Author List

Organisation	Name	Contact Information
UBI	Konstantinos Oikonomou	koikonomou@ubitech.eu
	Giannis Ledakis	gledakis@ubitech.eu
TRUST-IT	Marieke Willems	m.willems@trust-itservices.com
	Marialetizia Mari	l.mari@trust-itservices.com
	Zachary Smith	z.smith@trust-itservices.com
	Julie Abergas-Arteza	j.arteza@trust-itservices.com
ICCS	Panayiotis Michael	panayiotismichael@mail.ntua.gr
ICTLC	Martim Taborda Barata	martim.tabordabarata@ictlc.com
	Alberto Bettiol	alberto.bettiol@ictlc.com

Support & Review

Organisation	Name	Contact Information
DCU	Ray Walshe	ray.walshe@dcu.ie
	Xiao Rui Zhang	informationxiaorui.zhang@adaptcentre.ie

Abbreviations and Acronyms

ABAC	Attribute Based Access Control
AI	Artificial Intelligence
CERIF	Common European Research Information Format
CRS	Common Reporting Standard
FOAF	Friend Of A Friend
GDPR	General Data Protection Regulation
HL7	Health Level 7
ISO	International Organisation for Standardization
SI	Subjected Ingestion
SIOC	Socially Interconnected Online Communities
SMS	Service Management System
SSN	Semantic Sensor Networks

Table of contents

1. Introduction 4
Acknowledgements 4
2. Landscape analysis of Data Driven Policymaking 5
3. Policy Cloud multilayer framework for Interoperability 6
3.1 3-Dimensional end-to-end analysis using a novel Interim Repository Workplace 6
3.2 Building legal and ethical requirements into the analytics function registration process 8
4. On the need for a Policy Model standard 10
5. Conclusion & next steps 11
References 12

1. Introduction

Policymaking has always revolved around the initiative to solve problems and since the start of more focused research on it, it contained many stages from problem identification and agenda setting to policy formulation, legitimization, implementation and evaluation. As policymaking evolved and societal challenges became more complex, policymakers tried to include more factors into the decision-making process and thus data driven policymaking became an important tool in this field. Data driven policymaking aims to make optimal use of data and extensive analysis in order to create and evaluate policies.

The handling of data that is required has raised various legal and ethical issues that are hard to tackle on a case to case basis, without solid guidelines and standards. This need for standardization in data driven policymaking is more crucial due to the heterogeneity of the modern data sources and the interoperability of data. Several standards have been developed around this notion and were examined in the scope of the H2020 Policy Cloud project.

In order to cover the project's needs, a 3-dimensional end-to-end analysis is proposed, which uses a novel Interim Repository to facilitate the ethical and legal analysis on the interoperable data. This approach tries to find the balance between legal/ethical compliance and registrability of analytics, a fine line that most policymaking services need to optimize.

Konstantinos Oikonomou

Acknowledgements

The development of this white paper was a collaborative effort that was aided by StandICT.eu 2023, which is an EU H2020 Coordination and Support Action with the central goal of ensuring a neutral, reputable, pragmatic and fair approach to support European and Associated states presence in the international ICT standardisation scene.

StandICT.eu 2023 organised the Technical Working Group (TWG) for this whitepaper, called Data Driven Policy (TWG DDP) under the supervision of Ray Walshe - StandICT.eu 2023 EUOS Chair.

2. Landscape analysis of Data Driven Policymaking

Among the main value propositions of the Policy Cloud environment and tools for policy development and management is its ability to integrate, link and unify the datasets from diverse sources, while at the same time enabling analytics over the unified datasets. As a key prerequisite to providing this added value, the interoperability of diverse datasets should be ensured. Currently, a wide array of data representation standards in various domains have emerged as a means of enabling data interoperability and data exchange between different systems. Prominent examples of such standards in different policy areas include:

- The *INSPIRE Data Specifications* [1] for the interoperability of spatial data sets and services, which specify common data models, code lists, map layers and additional metadata on the interoperability to be used when exchanging spatial datasets;
- The *Health Level 7 (HL7)* [2] electronic message format for the exchange of electronic health information, includes interfaces for accessing these data in a unified and interoperable manner.;
- The *Common European Research Information Format (CERIF)* [3] for representing research information and supporting research policies;
- Data sharing agreement frameworks, such as the ISO/IEC 23751 [3] standard for cloud computing and distributed platforms;
- *Internet of Things ontologies and schemas*, such as the W3C Semantic Sensor Networks (SSN) ontology [5] and data schemas developed by the Open Geospatial Consortium (e.g., SensorML) [6];
- The *Common Reporting Standard (CRS),* developed in response to the G20 request and approved by the OECD Council on 15 July 2014, which specifies guidelines for obtaining information from financial institutions and automatically exchanging that information in an interoperable way;
- Standards-based ontologies appropriate for describing social relationships between individuals or groups, such as the *Friend Of A Friend (FOAF) ontology* [7] and the *Socially Interconnected Online Communities (SIOC) ontology* [8]. The latter are instrumental for analysing social media information.

These standards provide the means for common representation of domain-specific datasets, which provide the means for data interoperability (including in several cases semantic interoperability) across diverse databases and datasets. Nevertheless, these standards are insufficient for delivering the promise of Policy Cloud for several reasons, including:

- *Lack of semantic interoperability in the given domain.* For example, compliance to ontologies about IoT and sensor data fails to ensure a unified modelling of physics and mathematics, which are at the core of any sensing task. Hence, in several cases, there is a need for extending existing models with capabilities for linking/relating various quantifiable and measurable (real-world) features to define, in a user understandable and machine-readable manner the processes behind single or combined tasks in the given domain;
- *Lack of semantic interoperability across datasets from different sectors.* There is no easy way to link related information elements stemming from datasets in different sectors, which typically comprise different schemas. For example, environmental datasets and transport datasets contain many related elements, which cannot however be automatically identified and processed by a system due to the lack of common semantics;
- *Lack of process interoperability.* Policy Cloud deals with data driven policy development and management, whih entails the simulation and validation of entire processes. Especially in the case of multi-sectoral considerations (e.g., interaction and trade-offs between different policies) process interoperability is required to assess the impact of one policy on another.

In section 3, the multi-layer framework for interoperability across diverse policy-related datasets, which will facilitate semantic interoperability across such datasets (both within a single sector or across different policy sectors) is described.

3. Policy Cloud multilayer framework for Interoperability

Ensuring the interoperability of data and setting up data sharing services between different sectors and domains while building on existing European [9], international, or national standards is one of the vital parts of the European Commission's data strategy [10].

To this end, the Policy Cloud project has sought to support the key role of the Commission and the Member States in fostering better cooperation across all levels of public administrations in the Union, especially by breaking down the organisational and digital silos. The project has accordingly focused on public administrations' need for more specific guidance on how to improve governance of their interoperability activities to establish cross-organizational relationships, streamline processes and support end-to-end digital services [11]. More specifically, Policy Cloud partners have worked on how to further improve end-to-end integration and automation, making better use of reliable sources of information and openly publishing public data, while ensuring that citizens' and businesses' records are treated in accordance with applicable legal and ethical requirements [12].

3.1 3-Dimensional end-to-end analysis using a novel Interim Repository Workplace

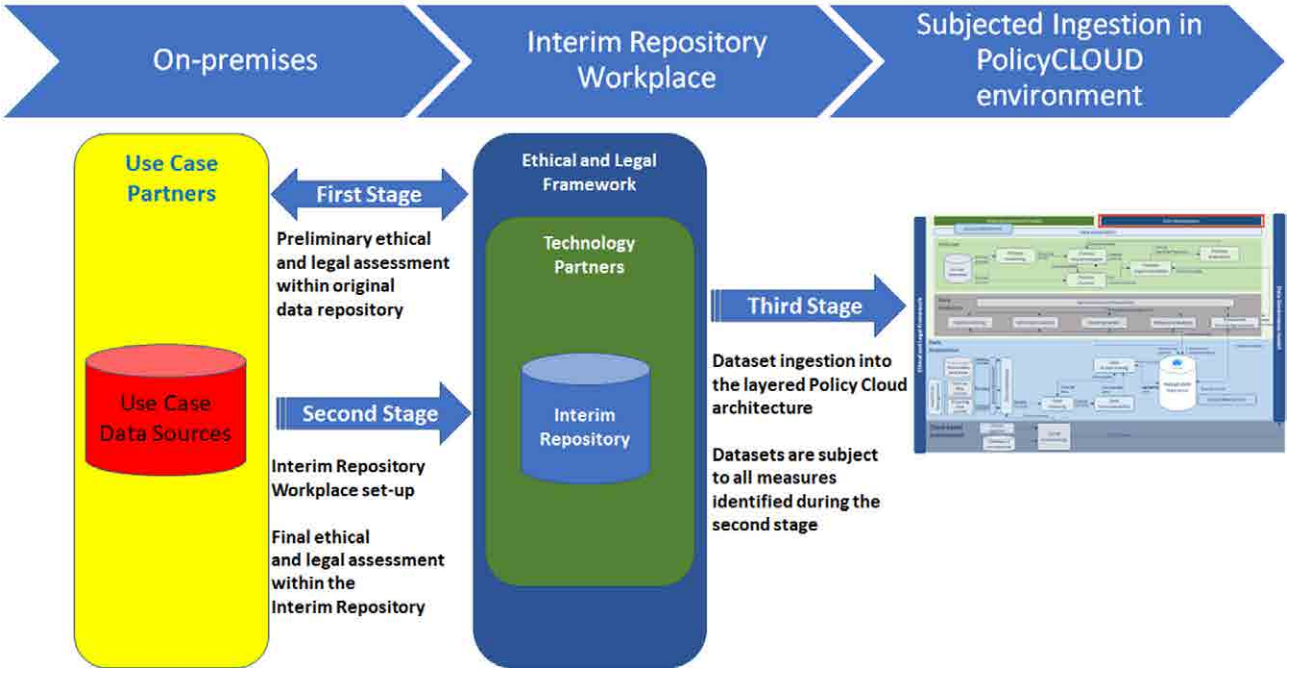


Figure 1: 3-dimensional analysis workflow

The efforts described in the previous section follow a three-stage approach (described below), known as an "**end-to-end legal and ethical assessment"**, which was developed jointly by a team consisting of the Policy Cloud Legal, Use Case and Technology partners. ***This three-stage approach can set the ground for standardising the related processes*** [13]**.** Aside from describing this approach below, we also include an illustration of how it may work in practice, based on assessments carried out in the scope of the Policy Cloud project.

The **first stage** involves a preliminary assessment of the implications around the use of available data within their original data repository/ies (e.g., publicly available website, or a site with authorised access) from a legal and ethical perspective without performing any data movement or processing. The goal of this preliminary assessment is to ensure that teams of legal and technology consultants, which have been authorised by the relevant data owner (e.g., a public administration), can access and assess the data in question in compliance with applicable laws and in an ethically sound manner.

For example, one of the scenarios tackled in the Policy Cloud project involves the use of signals received via

a call centre operated by the Municipality of Sofia (Bulgaria) related to road infrastructure and adjacent urban infrastructure within the city. This scenario seeks to provide a detailed analysis of the frequency and territorial distribution of these signals over time, based on various parameters (e.g., categories, types, areas, districts and major transport roads), with the overall goal of supporting and facilitating data-based municipal decision-making concerning the city's roads and infrastructure.

Two main initial concerns were identified: (1) this dataset would include personal data, notably information on citizens or other individuals raising signals, and (2) there may be contractual or intellectual property-related restrictions applicable to the use of this dataset for the intended purposes.

After a preliminary assessment, it was established that (1) a legal basis could potentially be identified under EU and local privacy/data protection law for the reuse of these personal data for the intended purposes, the collection of which could be supported by an information notice and minimised to the extent strictly necessary to allow the analytics to be successfully carried out (e.g., removing citizen/individual identifiers), (2) and (3) the Partner involved in the project which triggered this scenario is also legally responsible for management of the dataset, such that no issues around contractual/intellectual property-related restrictions should arise. Given the overall positive result, it was possible to move to the second stage of this approach.

In the **second stage**, a technology workplace is set up, called the Interim Repository. Provided that the preliminary assessment carried out over a given dataset is positive (i.e., the possibility to access and assess the corresponding data in a legally and ethically sound manner is confirmed), this dataset will be uploaded to the Interim Repository. Within this second stage, a final assessment of the legal and ethical soundness of the use of the given dataset for one or more specified purposes is then carried out. Where potential legal/ethical constraints are identified (e.g., the dataset contains personal data which is not necessary for the analysis purpose which is intended), specific technical/organisational measures are identified and implemented to deal with each constraint (e.g., the dataset is to be adequately cleaned of unnecessary personal data before further analysis in the third stage, to better protect the privacy and data protection rights of any individuals whose data may be included therein). In this sense, the **second stage** comes before the sophisticated process of fully ingesting data into the cloud environment and aims to:

1. Remove data/information silos;
2. Provide an introductory stage in enabling data interoperability;
3. Provide a legal and ethical "umbrella" for the data stored in the repository;
4. Provide safe and authorised access to data owned by Use Case partners;
5. Provide datasets that will be used for cross-domain evidence-based policymaking; and
6. Provide datasets that will be used for replication through the learning of Policy Makers from best practices applied to other use cases on a common (horizontal) domain of interest (e.g., environmental protection).

Specifically for safe and authorised access, Policy Cloud has prioritised this as a pillar of security and as a platform that handles user data. The Interim Repository is secured both via a user authentication mechanism called Keycloak, as well as an access control mechanism called Attribute Based Access Control (ABAC). Keycloak ensures that only users with valid credentials can interact with the Interim Repository while ABAC limits access to users that conform with the attributes defined in the implemented Access Policy. Both mechanisms working in conjunction ensure the security of the Interim Repository and authorise the source of any manipulation to the data. The same strong securities of the Interim Repository are extended to the whole Policy Cloud environment, as all components utilise the two aforementioned mechanisms.

After having uploaded the relevant dataset to the Interim Repository, further assessments were carried out as to the potential privacy/data protection implications around the use of the dataset (given that the initially identified contractual/intellectual property issues were addressed in the preliminary assessment). Examples of recommendations arrived at include the following:

1. *The relevant Partner would need to ensure that it has a legal basis to process personal data (and potential special categories of personal data) under the GDPR and under Bulgarian privacy/data protection laws - given that this processing was to be carried out for statistical purposes, its reuse was found to be admissible under Art. 25m of the Bulgarian Personal Data Protection Act;*
2. *An information notice was to be developed to ensure that citizens/individuals who had raised signals in the past, and which were to raise signals in the future, would be aware of this use of their signals and, as a result, able to exercise their data subject rights under the GDPR and Bulgarian privacy/data protection laws - in particular, the information notice published on the Municipality's relevant website was proposed to be updated to cover these activities;*

3. *The dataset was subjected to close scrutiny to confirm (1) whether it would be possible to proceed with the analytics activities with only anonymous or aggregated information, and (2) if not, what would be the absolute minimum amount of data needed to carry out these activities. It was determined that the dataset would be purged of identifiers concerning the signal submitters, and that even though the form used to submit signals includes free text fields (which might include personal data not easily purgeable), these analytics activities would not include any textual analysis of such fields.*

Given the overall successful implementation of recommendations proposed, the dataset was confirmed as ready for release (i.e., the third stage of this approach).

Finally, during the **third stage**, the dataset (subject to any measures identified during the **second stage**) is "released" from the Interim Repository and made available for further analysis.

The **difference between the second and third stages** is the fact that datasets during the second stage are visible by collaborators, and they are not hidden by the complex engineering mechanisms of the third stage, which are required for a cloud environment to perform. Still, during the second stage, discussions and agreements at political and legal levels are performed, guiding the technology steps that will take place at the third stage, which includes methods such as data anonymization and aggregation.

The main legal and ethical criteria on which the above described assessment activities are based are those related to data protection (with the General Data Protection Regulation - (EU) Regulation 2016/679 being the most important reference to this regard) and intellectual property and database protection laws.

However, to ensure ethical, regulatory, and societal sustainability of the project, not only statutory regulations are relevant, but also voluntary standards, such as:

1. IISO/IEC 27013:2021, on the integrated implementation of an information security management system (ISMS), as specified in ISO/IEC 27001 and a service management system (SMS), as specified in ISO/IEC 20000-1;
2. The CSA Cloud Controls Matrix, which is a cybersecurity control framework for cloud computing developed by the Cloud Security Alliance (currently on Version 4);
3. The CSA Code of Conduct for GDPR Compliance, which is a code of conduct under Art. 40 GDPR (currently undergoing an approval process with the French Data Protection Authority – the CNIL), designed to provide a consistent and comprehensive framework for cloud service providers to comply with the GDPR;
4. ENISA's European Cybersecurity Certification Scheme for Cloud Services (still in a draft stage at present);
5. EU Cloud Code of Conduct. It consists of requirements for Cloud Service Providers that wish to adhere to the code, plus a governance section that is designed to support the effective and transparent implementation, management, and evolution of the code. The code is a voluntary instrument, allowing Cloud Service Providers to evaluate and demonstrate its adherence to the code's requirements, either through self-evaluation and self-declaration of compliance and/or through third-party certification. The code has been developed to cover GDPR requirements and, following the positive opinion issued by the European Data Protection Board, has been officially approved by the Belgian Data Protection Authority in May 2021.

3.2 Building legal and ethical requirements into the analytics function registration process

Given that the Policy Cloud platform has been designed to allow for analytics tools to be registered for use via the platform, such tools should accordingly be held to a high standard of legal/ethical compliance. This is important, *inter alia*, to ensure that the platform can remain lawfully useable in the EU (e.g., by preventing the registration/use of tools which do not meet applicable legal requirements), to ensure the platform's data security (e.g., by preventing the registration/use of tools which may compromise the platform's integrity, or the confidentiality of data stored on the platform), to preserve the platform's trustworthiness (e.g., by preventing the registration/use of tools which do not meet baseline ethical standards, and which thereby present a relevant risk of deriving skewed, biased, inaccurate or otherwise misleading information from the data source to which they are applied) and to prevent potential reputational damages suffered by the platform managers. As such, technical/organisational measures to provide assurances in this respect must also be implemented on the platform, in particular to ensure that the following risks were reasonably and substantially minimised:

- Output generated by these activities leading to skewed, biased, inaccurate or otherwise misleading information being derived from a given data source, which may potentially culminate in misguided policy-making activities;
- Opacity in the output generation process, such that the ability to explain/understand how a given output was generated by these activities from a given input (as well as the possibility of false positives/negatives or other errors) is reduced or non-existent, which may potentially prevent or disincentivise Policy Cloud users from critically examining the output / information derived from a given data source during their decision-making process.

As AI-based technologies (including machine learning) are leveraged in connection with some of these tools, adherence to the 7 key requirements described in the AI-HLEG's Ethics Guidelines for Trustworthy AI [14] was also recommended as an important step towards maximising the trustworthiness of the Policy Cloud platform.

A balance must, however, be struck between maximising legal/ethical compliance and maximising the registrability of analytics tools. Overemphasis on compliance may create an overly restrictive registration process for such tools, which may ultimately compromise the effectiveness of the Policy Cloud platform in allowing tools other than those pre-existing to be leveraged; however, overemphasis on registrability triggers all risks related to the failure to meet a high standard of legal/ethical compliance mentioned above.

A practical initial answer found to this dilemma was the requirement for Policy Cloud users seeking to register a tool to document measures taken to address applicable legal/ethical requirements, through adequate fields added to the registration Application Programming Interfaces (APIs). In particular, specific input parameters to be addressed by tool registrants include requirements to provide details on specific measures taken to address the risk of biases inherent to the functioning of the tool, and on the relevant trade-offs encountered in the development of the tool, decisions made concerning the balancing of competing requirements (e.g., result precision vs. fairness) and measures taken to implement and document those decisions. This input can then be linked, on the Policy Cloud platform, to the analytics tool upon its successful registration. This can later be accessed by any Policy Cloud user wishing to make use of such a tool.

This answer provides some degree of assurance to users regarding the legal/ethical soundness of a given tool, and allows users to (1) make more informed and risk-based decisions about whether or not to leverage a given tool, to (2) provide information to relevant stakeholders about the tool if leveraged, and (3) to critically examine the output generated by the tool in the context of their policy-making decisions. However, for the future, it would be relevant to develop some standard form of communication of this information on different analytics tools which may be registered on the platform (and on other similar platforms), to facilitate user comparisons between tools and to allow greater interoperability with other systems performing similar or related activities.

4. On the need for a Policy Model standard

In addition to the need of finding and using common standards for data production and management within the EU, there is a need to create and use common standards for policymaking based on the use of information derived from these data. This dimension has not yet received the attention it deserves, but it is necessary to record the long journey from data management to reality-based decision-making.

The concept of the policy model is central to this transition. Its usefulness has been demonstrated in the previous Horizon 2020 project CrowdHEALTH [15] where it referred to Health policies [16] and has been extended to Policy Cloud for policies of any scope. Standardising the policy model can lower the communication barrier to policy decision-making by speeding up processes such as comparing policy-making processes, finding similar approaches between European Community countries, comparing effectiveness and versioning during their implementation. The proposed policy model of Policy Cloud, simple in its basic design but with the option of adding additional semantic levels, can provide a "fingerprint" of the purpose, need, stakeholders and monitoring indicators that lead to a policy decision.

Additionally, although it is essential to standardise the structure of a policy model, so that there is a common language in communication between policymakers, and in their accountability to the public, this effort should go even further. The results from the use of Analytical Tools to calculate the Key Performance Indicators (KPIs) that support a policy are also layers of information that also need to be standardised. Each analytical tool that in a given period processes specific data to produce information in the formulation and monitoring of a policy, creates results that are components of a new object (Analytics Result Object). AROs are the building blocks for calculating KPIs, but they are suitable as meta-data for input into other Analytical Tools as well.

The standardisation of the policy model in the first phase and the analytic results afterwards, will give an unprecedented momentum to the use of decision support systems for the production and monitoring of policies based on actual data and relative indicators.

5. Conclusion & next steps

Standardization of data interoperability in data driven policymaking is a topic that can affect a wide variety of technological, ethical and legal cases and is becoming more important as data sources become more readily available and diverse. Some standards have been already established in this field but further effort is needed to ensure legal and ethical compliance in more complex and heterogeneous cases. This need was obvious during the development of the Policy Cloud project and as part of it a 3-dimensional approach for an end-to-end analysis has been developed. This approach utilized a novel Interim Repository and proved efficient in the context of the project.

The next steps for this approach and for standardization in this field is to explore solutions that will maintain the fine balance between maximizing legal/ethical compliance, which can create over restrictive registration and onboarding processes and maximizing registrability of data analytics tools, which in turn can lead to failure to meet legal/ethical standards. The proposed approach of Policy Cloud to require assurance of legal/ethical measurements and documentation upon user registration led to increased awareness and soundness of legal/ethical issues in data analytics but it can be further improved and fine-tuned to achieve even greater balance between compliance and registrability.

References

[1]	EC INSPIRE KNOWLEDGE BASE, Data Specifications https://inspire.ec.europa.eu/data-specifications/2892, retrieved 2021-12-20.
[2]	HL7, project homepage, http://www.hl7.org, retrieved 2021-12-20.
[3]	CORDIS, Cerif http://cordis.europa.eu/cerif, retrieved 2021-12-20.
[4]	ISO/IEC 23751:2022, Information technology — Cloud computing and distributed platforms — Data sharing agreement (DSA) framework.
[5]	W3, Incubator https://www.w3.org/2005/Incubator/ssn/ssnx/ssn, retrieved 2021-12-20.
[6]	OPEN GEOSPATIAL, Standards http://www.opengeospatial.org/standards/sensorml, retrieved 2021-12-20.
[7]	FOAF, project homepage, http://www.foaf-project.org, retrieved 2021-12-20.
[8]	SIOC, project homepage http://sioc-project.org, retrieved 2021-12-20.
[9]	EUROPA.EU, Strategy for Data | Shaping Europe's digital future, https://digital-strategy.ec.europa.eu/en/policies/strategy-data, retrieved 2021-12-20.
[10]	COM(2020) 767, "Proposal for a Regulation of the European Parliament and of the Council on European data governance".
[11]	COM (2017) 134, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. European Interoperability Framework – Implementation Strategy".
[12]	Policy Cloud D3.6 Societal and Ethical Requirements and Guidelines, https://zenodo.org/record/5970753.
[13]	Policy Cloud D7.7, Standardisation Plan and Activities, https://zenodo.org/record/5948676#.YfqKu-rMI2w.
[14]	High-Level Expert Group on Artificial Intelligence, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, retrieved 2022-09-23.
[15]	K. Moutselos, D. Kyriazis and I. Maglogiannis, "A Web Based Modular Environment for Assisting Health Policy Making Utilizing Big Data Analytics" in 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), 2018.
[16]	Moutselos K, Maglogiannis I. Evidence-based Public Health Policy Models Development and Evaluation using Big Data Analytics and Web Technologies, 2020.

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

The StandICT.eu 2023 project has received funding from the European Union's Horizon 2020 - Research and Innovation programme - under grant agreement no. 951972.