"Building the Next Generation Personal Data Platforms"
G.A. n. 871370

DELIVERABLE D9.2
POPD - Requirement No. 2

H2020-EU: **PIMCity**
Project No. 871370
Start date of project: 01/12/2019
Duration: 30 months

Revision:
Deliverable delivery date: 30/06/2020
Deliverable due date: 31/05/2020

Document Information

Document Name: Data Management Plan - Mitigation actions POPD - Requirement No. 2
WP9 – Ethics requirements
Task ===
Revision: V01
Revision Date: 16/06/2020
Author: POLITO and all Partners

Dissemination Level

Project co-funded by the EC within the H2020 Programme		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	☑

(Tick the corresponding dissemination level of the deliverable according to Annex I).

Approvals

	Name	Entity	Date	Visa
Author	WP8 Team	POLITO	11/06/2020	
WP Leader	Marco Mellia	POLITO	16/06/2020	☑
Coordinator	Marco Mellia	POLITO	16/06/2020	☑

Document history

Revision	Date	Modification
Version 1	16/06/2020	V1

List of abbreviations and acronyms

Abbreviation	Meaning
G.A.	**Grant Agreement**
CA	**Consortium Agreement**
GA	**General Assembly**
PB	**Project Board**
PC	**Project Coordinator**
PrO	**Project Office**
IR	**Interim Reports**

Disclaimer

Index

INTRODUCTION

In order to share information inside Consortium, raise awareness about the topic, provide/share advices, collect information conf-calls among Partners have been organized within WP7 and WP9 on:

- Dec 19th 2019 organized by POLITO with all Partners
- Jan 20th 2020, organized by POLITO with all Partners
- Feb 17th 2020, organized by POLITO with all Partners
- Apr 8th 2020, organized by POLITO with all Partners
- Apr 22nd 2020, organized by POLITO with all Partners
- May 8th 2020, organized by POLITO with all Partners
- May 27th 2020, organized by POLITO with KUL in order to discuss D9.1/D9.2 content and D7.1/D7.2 content.

In general, the research activities in the project foresee the:

- ✓ Automatic processing of web page content to generate privacy tags that summarize the usage of personal data of a website
- ✓ Automatic processing data aggregation from advertising platforms and open rtb platforms to obtain a large-scale database of audience values from market-side perspective
- ✓ Automatic collection and processing of data from websites using web scrapers to generate Privacy Metrics, i.e., sets of scores and metrics to allow users understand which personal data is collected and how from websites and third parties connected to them
- ✓ Automatic calculation related to the value that data is bringing to a certain machine learning, targeting or prediction model
- ✓ Process implementation to allow users to migrate their data to new platforms, in a privacy-preserving fashion (*data portability procedures*)
- ✓ Collection of aggregated network traffic data like number of users/sessions per public website or service, or bandwidth usage per website/service
- ✓ Aggregation and index data collected from Personal consent manager (P-CM) to provide API to other components from the project. Trading engine (TE) will execute contracts based on buyers' queries and user preference.

As a general point of view the research project plan is to use publicly available datasets, complemented by automatic collection of website data using scrapers, and to install in potentially thousands of users a browser plugin that capture the hostnames visited by the users. Only the latter constitutes a potential collection and processing of personal data. That data will be then analyzed to generate user profiles that will be traded with third parties, all with the consent of the user providing the data.

In the first phase of the project, the public data will be stored at NEC premises (ANT testbed), only providing info to the users through an interface. In the final version (the whole PIMS implementation, with data being traded) the data will be stored, analyzed and possible traded in a cloud system provided by Fastweb. This cloud is to be defined in the second year of the project.

Finally, it is forecasted the participation of voluntary end-users to fill in surveys and to test PIMCity applications as beta testers through the web www.pimcity.eu.

Please note that the following content refers to D7.1, D7.2 and to the DPIA done by each Partner.

1. DATA MINIMIZATION PRINCIPLE AND "PRIVACY BY DESIGN"

"*Privacy by design*" and data minimization requirements are documented in deliverable (D7.2), which is disseminated among partners to provide necessary information that should be taken into account from the begin of the process.

In other words, from a legal perspective, organizational measures have been focused on defining and informing all partners about the key aspects and requirements stemming from the EU data protection framework.

Besides, KUL (as legal Partner) has provided guidelines for informed consent forms, templates for informed consent, guidelines to identify the controller, joint controllers and processor, guidelines for privacy policies and processing of data in the online context, including a template for privacy policy.

The DMP detailed in D7.1 is considered a live document and it will be updated at the later stage of the project when some crucial details of data processing will be detailed more in deep.

2. PROCEDURE TO MITIGATE POTENTIAL "REVERSE ENGINEERING", SPECIFICALLY IN THE CONTEXT OF DATA MERGING

A first overview is defined in D7.1 and 7.2.

Specific requirements will be detailed at a later project stage as, at the moment of writing (M6), some crucial details of data processing need to be defined more in depth.

3. PROCEDURE TO AVOID UNFORESEEN IDENTIFIER DISCLOSURE AND HOW TO PROPERLY MANAGE THE PROCESS OF DATA COLLECTION

According to the DPIA done by Partners, the procedure of data collection and storage will include these protection measures against involuntarily disclosure of personal data:

- ✓ Network security measures in partners premises including hardware security (firewall, traffic monitoring, server access control through users and passwords)
- ✓ Physical access control to datacenters hosting databases
- ✓ Logical access control to databases
- ✓ Usage of cloud computing infrastructure located in Italy and EU only

In details, all data will be stored in servers which are protected with start-of-the-art security techniques implemented by the *IT & Security department* including firewall, traffic monitoring, server access control through users and Password management. Servers are located in rooms with physical access control, backups, disaster recovery mechanisms, etc.

Data will be stored in servers and containerized servers, which are protected with *start-of-the-art security techniques and technologies as well*

Whenever possible, all data will be encrypted using *state-of-the-art encryption technology and techniques*. The system will optionally provide anonymization functions for preserving the user's privacy, based on the type of data that will be processed. More sophisticated mechanisms, such as differential privacy, could also be available.

Data integrity will be guaranteed with cryptography techniques to ensure restricted access to backups. Data will be periodically backed up. All the access to the servers will be logged and kept for security audits.

In some cases, the servers where the data is stored will not have direct access from the public Internet. Servers are only physically accessible to authorized personnel.

Considering data collection and movement, all the data will be transferred using encrypted connections.

For PII – we will consider the opportunity to store PII differently from data subjects' identifiers. If available, the PII like the name of the user will not be stored in the same premises as for instance the browsing data collected.

On a general point of view, as Consortium, there is not any standard applicable to the processing of data being this a novel approach. In general, best practices will be adopted by each single Partner: data is maintained encrypted and transferred using secure channels. Data is accessible and processed by authorized personnel only, possibly using automatic procedures.

4. PROCEDURE TO AVOID DATA OVER-COLLECTION, INCLUDING THE COLLECTION OF SENSITIVE DATA AND EXPLAIN HOW DATA WILL BE COLLECTED ONLY ON A STRICT "NEED TO KNOW" BASIS

According to *EU Grants: Horizon 2020 Guidance —How to complete your ethics self-assessment: V6.1–04.02.2019*, the following main obligations will be followed:

- ✓ Data processing is subject to appropriate safeguards (see above)
- ✓ Data is wherever possible processed in anonymized or pseudonymized form
- ✓ Data processing is subject to free and fully informed consent of the persons concerned (unless already covered by another legal basis, e.g. legitimate or public interest)
- ✓ Data processing is NOT performed in secret. Research participants must be made aware that they take part in a research project and be informed of their rights and the potential risks that the data processing may bring.
- ✓ Data may be processed ONLY if it is really adequate, relevant and limited to what is necessary for your research (*'data minimization principle*').

The level of data security will be appropriate to the risks for the research participants occurring in case of unauthorized access or disclosure, accidental deletion or destruction of the data.

The Consortium is responsible for any partners, contractors or service providers that process research data at your request or on its behalf. Generally, one of the best ways how to avoid/limit data protection issues for your project is to use anonymized or pseudonymized data.

As indicated in DPIA, for the purposes of the project, Partners will collect data as:

- ✓ aggregated data from advertising platforms and open RTD platforms about the value of audiences
- ✓ aggregated anonymized mobility data from the Internet to feed the data valuation algorithms
- ✓ web pages by automatically crawling websites
- ✓ web pages visited by different users with a browser plugin; that data will be used to generate user profiles. This kind of data can include sensitive data such as sexual orientation (i.e., if the user visits homosexual dating websites) or religious facts (i.e., if the user visits specific websites with religious content), etc.
- ✓ Other limited data could come from Social Media - Mobile Operator Data – eventually Banking, Emails, Calendar.

Moreover, data will be collected through forms on the website. Once data arrives at Partners' web server, the identification data is stored in encrypted form in a table and the rest of the data is disaggregated and stored in order to be able to work with it for the purpose of making calculations and statistics.

If data will consist on tuples <*identifier_of_user_data, consent_level*> for each user, the consent_level is a structure common across all users. The *identider_of_user_data* is a pointer to arbitrary user's data stored for instance on the P-DS. On the scope of the Trading Engine (TE), the data will consist on contract, queries expressed by buyers. In the case of fulfillment of the contract, it will be stored for auditing purposed. Contracts contain no personal data but pointers to all relevant actors of it, namely: *identider_of_user_data, identifier_of_seller, intefider_of_buyer* and other metadata relevant for the contract such as date, expiration, price, etc.

Partner's web scrapers download all files needed to render the page (e.g., HTML, images, CSS, and Java script files) as well as logs to API calls executed by the browser when rendering the page.

All data will be temporarily saved until the user decides to empty the browser storage. Only the user that owns the data will have access to data.

Please note that the goal of the data collection and processing follows the data minimization collection principle; PIMCity Consortium highlights that the data collection purpose is (as an explicit and legitimate goal in the context of the project execution):

- ✓ To provide privacy tags
- ✓ To provide a third party estimation about the value of different data sources in a certain prediction task, and derive an algorithm that would be useful in a loosely defined context
- ✓ To provide third party estimation about the value of different audiences
- ✓ To generate Privacy Metrics, and consequently provide the users with tools to understand privacy risks connected to given web services. Data are not used for other purposes
- ✓ to be able to contact users who want to contribute to carrying out surveys or testing the results aimed at users
- ✓ to realize the data portability functionality of the PIMCity project
- ✓ (for P-CM) is to let the users control their data and to provide API access for other components and for TE is to execute the contract and for later auditing. The information will be presented to the user explicitly either as part of consent or in the Terms and Conditions

On the other hand, taking into account the recent EDPB's opinion (EDPB' Guidelines 05/2020 on consent under Regulation 2016/679 Version 1.1 Adopted on 4 May 2020) on the notion of consent and the case law of the Court of Justice of the European Union, partners are currently discussing which medium should be used during the process to gain data subjects' consent. At this stage of the project partners have not agreed from which source data will be collected.

5. PROCEDURE ON HOW THE CONSORTIUM INTENDS TO ASSESS AND CONSIDER THE PROJECT DATA RELATED INCIDENTAL FINDINGS

A first overview is defined in D7.1 and 7.2.

Specific procedures will be detailed according to a later project stage, as at the moment (M6), some crucial details of data processing need to be defined more in depth.

6. CONCLUSION

The DMP, as indicated and described in D7.1, provides that consent will be collected taking into account all relevant legal requirements. The project will follow all the current practice to protect any eventual PII and sensible pieces of data in accordance to the data minimization and privacy by design approaches. Similarly, all state-of-art technologies will be adopted to protect data from unforeseen and unwilling disclosure of data due to data breaches.

All of data will be processed in accordance with the relevant EU legal requirements. To enable data subjects to exercise their rights effectively, data controllers will fulfill all the requirements stemming from the relevant legal frameworks as defined in the deliverables of WP7, including the implementation of internal organizational and technical measures.

We stress that this document will be updated whenever needed to cope with new requirements that would arise in the future.