



# **DELIVERABLE D1.5**

#### **Demonstrators and evaluation report**

H2020-EU-2.1.1: PIMCity Project No. 871370 Start date of project: 01-12-2019 Duration: 33 months

Revision: 01

Date: 31-08-2022





#### **Document Information**

**Document Name: Demonstrators and evaluation report** 

WP1 – Title: Platform specification and demonstration

Task 1.2, 1.3 and 1.4

Revision: 03

#### **Revision Date:**

Authors: Roberto González Sánchez (NEC) as main editor, and all partners involved in WP1.

#### **Dissemination Level**

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

#### **Approvals and authors**

	Name	Entity	Date	Visa
Author	Roberto González	NEC	31/08/22	
Author	Roberto Bifulco	NEC	31/08/22	
Author	Daniel Oñoro	NEC	31/08/22	
WP3 Leader	Rubén Cuevas	UC3M	31/08/22	
Author	Angel Cuevas	UC3M	31/08/22	
Author	Martino Trevisan	POLITO	31/08/22	
Author	Marco Mellia	POLITO	31/08/22	
Author	Michela Meo	POLITO	31/08/22	
Author	Stefano Traverso	ERMES	31/08/22	
Author	Luca Moretto	ERMES	31/08/22	
Author	Davide Pozza	ERMES	31/08/22	
Author	Michele Massese	ERMES	31/08/22	
Author	Daniele Milani	ERMES	31/08/22	
WP2 Leader	Martino Trevisan	POLITO	31/08/22	
WP1 Leader	Roberto González	NEC	31/08/22	
Coordinator	Roberto González	NEC	31/08/22	
Project	Marco Mellia	POLITO	31/08/22	
Coordinator				
WP4 Leader	Nicolas Kourtellis	TID	31/08/22	





Author	Kleomenis Katevas	TID	31/08/22
WP5 Leader	Stefano Traverso	ERMES	31/08/22
Author	Daniel Fernandez	Wibson	31/08/22
Author	Rodrigo Irarrazaval	Wibson	31/08/22
Author	Miguel Pérez Subías	AUI	31/08/22

# **Document history**

Revision	Date	Modification
Version 1	15-07-2022	First ToC
Version 2	01-08-2022	First complete draft
Version 3	31-08-2022	Final version with final results





# List of abbreviations and acronyms

Abbreviation	Meaning	
PIMS	Personal Information Manager System	
PDK	PIMS Development Kit	
TT	Transparency Tags	
PDA	Personal Data Avatar	
UD	User Dashboard	
P-DS	Personal Data Safe	
P-PM	Personal Privacy Metrics	
P-CM	Personal Consent Manager	
P-PPA	Personal Privacy Preserving Analytics	
D-VT	Data Valuation Tools	
D-TE	Data Trading Engine	
DKE	Data knowledge Extraction	
NEC	Nec Laboratories Europe Gmbh	
FW	Fastweb spa	
POLITO	Politecnico di Torino	
TID	Telefonica Investigación y Desarrollo sa	
ERMES	Ermes Cyber Security S.R.L.	





# Executive Summary

D1.5 *Demonstrators and evaluation report* presents the final evaluation of the technology developed by the PIMCity project. The evaluation is done by performing three independent pilots designed to assess different parts of the solutions developed during the project.

The first pilot focuses on the evaluation of the different PDK components in a real environment. This evaluation is performed by stressing the different components using the data and platforms of the different SMEs participating in the project. This document presents an updated version of the content of D1.4.

The second pilot demonstrates the usability of the solutions for companies in a B2B setting. In this pilot, instead of independent PDK components, an integrated solution including the Transparency Tags, together with other products of ERMES is deployed in external companies.

The third pilot demonstrates the EasyPIMS platform. It demonstrates that by integrating the different components developed during the project it is possible do create a fully-fledged PIMS platform. During this pilot, we evaluate the interaction of the users with the platform, allowing them to share their data with third parties. Those third parties are companies willing to buy data from the platform. Moreover, we demonstrate the feasibility on the integration with telco companies.

The three pilots complement each other on the assessment of the PIMCity technology, either in a standalone way, helping other companies to improve their way of operating or integrated as a complete platform.





# Index

	Index	κ	6
1	Intr	oduction	8
_			
2	Con	tinuous Testing	10
	2.1	Design	10
	2.2	User Involvement	11
	2.3	Data Buyer Involvement	
	2 /	Evolution	11
	<b>2.4</b>	Execution Tools	11
	2.4.1	Execution Environment	11 11
	2.4.2		
	2.5	Results	12
	2.5.1	WP2	12
	Pe	ersonal Data Safe	12
	Pe	ersonal Consent Manager	13
	Pr	ivacy Metrics	13
	2.5.2	WP3	14
	Da	ata Valuation Tools from Market Perspective	14
	Da	ata Trading Engine	14
	2.5.3	WP4	15
	Da	ata Provenance	15
	2.5.4	WP5	15
	Та	isk Manager	15
	2.6	Evaluation and Recommendations	16
	2.7	Conclusions on the demonstrator	16
3	B2B	Testing	
	2 1	Domonstrator	10
	<b>5.1</b>	Demonstrator	10
	3.1.1		18
	3.1.2		
	3.1.3	Demonstration design	
	3.1.4	Demonstrator results	
	Q	1: What do you like of Transparency Tags?	21
	Q	2: What do you dislike of Transparency Tags?	22
	Q	3: What would you improve/expand?	
	Q	4: How much would you pay for TTs?	23
	3.1.5	Conclusions on the demonstrator	23
4	B2C	Testing	24
	4.1	Demonstrator design, the PDA	
	4.1.1	User engagement	26
	Pi	Iblic engagement and community building	
	Er	id-user participation	
	 Fa	svPIMS tasks	
		· · · · · · · · · ·	
		Personal data task	





C	Download Plugin Task	28
[	Display Ads Task	29
S	Set consents task	30
I	nvite friends task	31
ι	Jpload location history task	31
Adv	vertising	32
Swe	eepstakes	33
Oth	er communication actions	34
4.1.2	Demonstrator Evaluation	35
Tar	geted advertising	38
4.1.3	Users feedback	40
4.2 [	Demonstrator design, the Marketplace4	46
<b>4.2</b> [ 4.2.1	Demonstrator design, the Marketplace       4         Data Buyers involvement       4	<b>46</b> 47
<b>4.2 [</b> 4.2.1 Tria	Demonstrator design, the Marketplace	<b>46</b> 47 51
<b>4.2</b> [ 4.2.1 Tria 4.2.2	Demonstrator design, the Marketplace	<b>46</b> 47 51 51
<b>4.2 [</b> 4.2.1 Tria 4.2.2 4.2.3	Demonstrator design, the Marketplace	<b>46</b> 47 51 51 51
<b>4.2 [</b> 4.2.1 Tria 4.2.2 4.2.3 <b>4.3</b>	Demonstrator design, the Marketplace	46 47 51 51 51 51
<b>4.2 [</b> 4.2.1 Tria 4.2.2 4.2.3 <b>4.3 1</b> 4.3.1	Demonstrator design, the Marketplace       4         Data Buyers involvement       4         Is and testing       5         Demonstrator evaluation       5         Data buyers' feedback       5         Felco demonstrator       5         Browsing history demonstrator: Privacy preserving user profiling with Telco data       5	<b>46</b> 47 51 51 51 <b>55</b>
4.2 [ 4.2.1 Tria 4.2.2 4.2.3 4.2.3 4.3.1 4.3.1 4.3.2	Demonstrator design, the Marketplace       4         Data Buyers involvement       4         Ils and testing       5         Demonstrator evaluation       5         Data buyers' feedback       5         Felco demonstrator       5         Browsing history demonstrator: Privacy preserving user profiling with Telco data       5         Location History demonstrator: Location data valuation       5	<b>46</b> 47 51 51 51 <b>55</b> 59





# **1** Introduction

PIMCity is an industry-driven innovation action, and as such, it aims to testing the technical components developed in the relevant scenario (TRL-7 or higher) during the project. The PDK and the EasyPIMS platform have been independently piloted with real users to demonstrate both, the usage of the different components and the whole integrated platform.

#### **PDK DEMONSTRATION.**

The aim of the PDK is to produce modules that are easy to fit into existing and new PIMS. Two different pilots have been carried out to demonstrate how easy and useful the modules are. They target both a B2C and B2B cases.

#### Pilot 1: Continuous testing of the PDK components.

In the first pilot, the SMEs participating in the project have tested the different modules within their already existing customer bases following an Agile methodology. Similarly, to the continuous design-implementation-testing cycles typically adopted by ICT startups, the different components will be tested as soon as they are implemented.

The main goal of this pilot is to detect limitations and problems of the modules in a timely manner, providing feedback to improve their development. To this end, the SMEs selected the modules that fit better with their different applications.

In particular, the SMEs tested the main PDK components developed in WP2, WP3 and WP4, such as the Personal Data Safe, Personal Consent Manager and Data Trading Engine, and reported all the findings involving technical and functional unexpected limitations, as well as empirical Service Level Agreements (SLAs).

The initial results of this pilot were presented in D1.4. Section 2 of this document updates that results.

## Pilot 2: B2B testing of the PDK components

The second pilot is devoted to check the usability and versatility of the different components in the B2B operation of ERMES. To this end, FW and ERMES have jointly explore their customer portfolio to identify companies interested in the transparent trading of personal data. The main goal of this pilot is to demonstrate the B2B value of some components as well as the ease to generate complex solutions in different markets by integrating them. To this end, ERMES have develop solutions for companies that will help them protect their data and increase privacy and security awareness too.

Contrary to the previous pilot, that integrates the main PDK components developed in WP2, WP3 and WP4, ERMES has used of the TTs and PDA solution designs developed in WP5 and further developed them to integrate them in ERMES' products. The final outcome of this pilot will be an updated exploitation plan for the B2B commercialization of the PIMCity technology that will be carried out by ERMES and FW after the project (see D6.7).

To test the quality of implementation, 10 companies selected amongst ERMES' clients were contacted and 6 accepted to participate in the pilot. For these, the Transparency Tags module integrated in ERMES' web protection product, *Ermes for Enterprise*, has been enabled and used to collect feedback to check the exploitation feasibility of the solution.





Section 3 presents the results obtained during this pilot.

#### EASYPIMS DEMONSTRATION.

PIMCity also demonstrates the potential of the components when integrated into the EasyPIMS platform, targeting a large scale B2C use case.

#### Pilot 3: B2C testing of EasyPIMS with real customers.

This third pilot is the main demonstrator of the project. The project bootstrapped the EasyPIMS platform, involving several thousands of real users. This pilot tests all EasyPIMS components, namely, the PDA, the TT, the UD and the Marketplace. Fed by data already available at the operators like the browsing history, the call and position logs of the customers, it will let the users incorporate new personal data (i.e., the gender, age, preferences, interests, etc.) via the UD. The PDA will let them control the data that will be offered on the marketplace, while the TT will let them take informed decisions.

The main goal of this main pilot is to demonstrate the potential of EasyPIMS by testing it with a critical mass of users, allowing the partners to create their own data platform for its commercial use after the project. The outcome of this pilot is a business and exploitation plan of the EasyPIMS platform (see D6.7) as well as the outreach and engagement strategy to be followed after the project to keep growing the mass of users.

This pilot was split in different phases to test different parts of the EasyPIMS platform and its integration into the telcos. In total, 3360 (as of the 29/8/2022) real users participated in the main EasyPIMS pilot by registering in the open platform (<u>https://EasyPIMS.com/</u>) and using it on a regular basis. On the other hand, 16 data buyers participated and evaluated the data marketplace offered by the platform.

In addition, to test the possible integration of the platform within the telcos (that could directly feed data into the platform) we have implemented two different smaller demonstrators using data directly collected from the telecommunication companies participating in the project. Different PDK components were tested with data coming from more than 450k real users.

Section 4 presents the results obtained during this pilot.





# 2 Continuous Testing

During the PIMCity project we performed three completely independent pilots designed for different purposes. This deliverable presents the design, the execution, and results of the Continuous Testing Pilot.

The Continuous Testing Pilot is designed to discover as soon as possible limitations and problems of the PDK components, allowing that way the correction of those problems before the technical WPs are finished following an agile programming approach.

On this pilot, the SMEs participating in the project (Wibson, ERMES, LSTech and TapTap) tested the main PDK components developed in WP2, WP3 and WP4, such as the Personal Data Safe, Personal Consent Manager and Data Trading Engine, and reported all the findings involving technical and functional unexpected limitations, as well as empirical Service Level Agreements (SLAs).

These results were included in the Deliverable 1.4 Agile Report and are now updated in this document with their corresponding evaluation, exposing the limitations and virtues of each PDK.

# 2.1 Design

Since each PDK was initially thought and designed to carry out a single responsibility, it seems natural to test them separately, under different conditions and circumstances, as well as involving different PIMCity partners. Nevertheless, the aim is always the same: at the end of the pilot, any possible limitation and problem within these PDKs must be reported to the developers to allow improvements during the duration of PIMCity.

The Continuous Testing pilot investigates the following measurements and metrics when applicable:

- Response KPIs:
  - Transaction average response time time to first byte or last byte
  - Peak response time tells the longest cycle
  - Error rate percentage of problems compared to all requests
- App and Network Load Testing
  - Throughput How many transactions per minute can the PDK API handle?
     When is it time to scale to more web server instances?
  - Bandwidth requirements Is the network a bottleneck? Or is there content that is pulling it down that we can offload?
- Resources Load Testing
  - CPU usage hotspots How much CPU usage do we need for each load?
     Can we fix programming to lower CPU or do we simply need more?
  - Memory problems Is there a memory leak?
  - Worker threads Are they correctly configured? Are there any web server modules that block these threads?





# 2.2 User Involvement

At this point, there is no need for real user involvement, as long as tests represent a real (or near real) scenario in terms of volume of data, interactions and expected API reactions. For example, it is not important if data is real or synthetic, but it must add up to hundreds of MB (at least) to represent an amount similar to an MVP running in a production environment.

In concrete numbers, the following volumes are used during this pilot, combining them in different ways to test different aspects of each PDK:

- Data volume: 1 to 100 MB
- Concurrent interactions: 10 to 100
- Transactions within a single session: 10 to 50

## 2.3 Data Buyer Involvement

Similar to User Involvement, there is no need for real Data Buyers at this point. Nonetheless, understanding their usual behaviour is key to carry out a satisfactory pilot. To these extent, Data Buyers are faked to buy high volumes of data, of different types, and within different timespans.

In concrete numbers, the following volumes are used during this pilot, combining them in different ways to test different aspects of each PDK:

- Data volume: 100 to 500 MB
- Concurrent interactions: 1 to 20
- Transactions within a single session: 1 to 10

## 2.4 Execution

#### 2.4.1 Execution Tools

The tools used to carry out the execution of the tests are Postman and the Swagger UI exposed by each PDK deployed in the production environment. Each test is executed at least three times during different moments to reduce fortuitous situations that might contaminate the results.

Postman<sup>1</sup> is an API platform for building and using APIs. Postman simplifies each step of the API lifecycle and streamlines collaboration so anyone can create better APIs—faster. Its main features are API repository, API client, Intelligence, Workspaces and Integrations.

Swagger UI allows anyone — be it a development team or end consumers — to visualize and interact with the API's resources without having any of the implementation logic in place. It's automatically generated from the OpenAPI definitions (formerly known as Swagger) included in WP5 deliverables, with the visual documentation making it easy for back-end implementation and client-side consumption.

# 2.4.2 Execution Environment

<sup>&</sup>lt;sup>1</sup> It can be downloaded for free from its main website <u>https://www.postman.com</u> in different versions and formats.





The execution environment is a vital part of the testing phase. With poor decisions, the tests might show that the solution is slow or not suitable for the project. That is why deciding on the resources for the environment takes time and a thorough analysis over the system requirements and expected operations. To build this report, all tests were run against each PDK's API hosted on virtual machines with the following characteristics:

Characteristic	Value
CPU Virtual Cores	8
RAM	32 GB
Storage	64 GB
Internet Connection	500 Mbit/s symmetric
OS	Linux Ubuntu 20.04.1 LTS

Each API was run using a Docker container with the same OS as the host server and same resources.

# 2.5 Results

The whole test suite defined in *Deliverable 1.4 Agile Report* was re-executed with the tools and environment here defined. In this section, the results obtained are exposed following their corresponding evaluation.

## 2.5.1 WP2

#### Personal Data Safe

ID	Result	Comments
G1	Avg. Response Time: 420ms	Keeps D1.4 efficiency.
G2	Peak Response Time: 968ms	Almost doubled D1.4 result. This is only for the case of testing with high volumes of browsing history data that were not used in previous report. Still meets the requirements.
G3	Max. Requests Per Minute: 15,200	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 6.12% Avg. RAM Usage: 8.71%	Keeps D1.4 efficiency.
PDS1	The record is inserted.	-
PDS2	The whole batch is inserted.	-
PDS3	The record is obtained.	-
PDS4	The record is modified.	-





PDS5	The record is deleted.	-
PDS6	The data uploaded by other users is obtained.	-
PDS7	The batch is inserted without issues.	-
PDS8	The records are inserted without issues.	-

# Personal Consent Manager

ID	Result	Comments
G1	Avg. Response Time: 302ms	Keeps D1.4 efficiency.
G2	Peak Response Time: 457ms	Keeps D1.4 efficiency.
G3	Max. Requests Per Minute: 15,160	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 9.45%	Keeps D1.4 efficiency.
	Avg. RAM Usage: 11.67%	
PCM1	The list is obtained.	-
PCM2	The consent is enabled.	-
PCM3	The consent is disabled.	-
PCM4	The list is obtained.	-
PCM5	The consent is enabled.	-
PCM6	The consent is disabled.	-
PCM7	The list of users with enabled consents is obtained.	-

# **Privacy Metrics**

ID	Result	Comments
G1	Avg. Response Time: 408ms	Keeps D1.4 efficiency.
G2	Peak Response Time: 743ms	Keeps D1.4 efficiency.
G3	Max. Requests Per Minute: 14,850	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 9.10%	Keeps D1.4 efficiency.
	Avg. RAM Usage: 7.98%	





PM1	The privacy metric for the buyer is created with its corresponding score and web data.	-
PM2	The list of all privacy metrics is obtained.	-
PM3	The privacy metric for that buyer is deleted.	-
PM4	The privacy metric for that buyer is obtained.	-
PM5	The privacy metric is updated.	-
PM6	The privacy metric is updated.	-
PM7	The privacy metric is updated.	-

# 2.5.2 WP3

# Data Valuation Tools from Market Perspective

ID	Result	Comments
G1	Avg. Response Time: 758ms	The average response time is higher than expected for any other PDK, but it is understandable in this case, since it needs to take the result from a third- party.
G2	Peak Response Time: 963ms	Although the average response time is high, the peak is within the acceptable boundary.
G3	Max. Requests Per Minute: -	This test was not done, since the maximum requests per minute allowed are determined by Facebook: https://developers.facebook.com/docs/graph-api/overview/rate-limiting/
G4	Avg. CPU Usage: 8.57% Avg. RAM Usage: 7.12%	Keeps D1.4 efficiency.
DVTMP1	The price is obtained.	-
DVTMP2	The price is obtained.	-
DVTMP3	The price is obtained.	-

## Data Trading Engine

31/08/2021





ID	Result	Comments
G1	Avg. Response Time: 433ms	Keeps D1.4 efficiency.
G2	Peak Response Time: 912ms	Keeps D1.4 efficiency.
G3	Max. Requests Per Minute: 15,120	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 11.43%	Keeps D1.4 efficiency.
	Avg. RAM Usage: 12.84%	
DTE1	The number of points is obtained.	-
DTE2	The number of credits is obtained.	-
DTE3	The number of credits spent is obtained.	-
DTE4	The credits are added.	-
DTE5	The price is obtained.	-
DTE6	The list of offers is obtained.	-
DTE7	The list of offers is obtained.	-
DTE8	The data is downloaded in a JSON- formatted file.	-

## 2.5.3 WP4

#### Data Provenance

ID	Result	Comments
G1	Avg. Response Time: 286ms	Keeps D1.4 efficiency.
G2	Peak Response Time: 530ms	Keeps D1.4 efficiency.
G3	Max. Requests Per Minute: 16,220	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 6.93% Avg. RAM Usage: 12.90%	Keeps D1.4 efficiency.
DP1	The URL is watermarked.	-

# 2.5.4 WP5

# Task Manager

ID	Result	Comments
G1	Avg. Response Time: 374ms	Keeps D1.4 efficiency.
	·	1





G2	Peak Response Time: 477ms	Keeps D1.4 efficiency.
G3	Max. Requests Per Minute: 15,560	Keeps D1.4 efficiency.
G4	Avg. CPU Usage: 7.64% Avg. RAM Usage: 12.72%	Keeps D1.4 efficiency.
TM1	The list is obtained.	-
TM2	The task is marked as completed, and the corresponding timestamp is specified.	-
ТМЗ	The task status belongs to the latest period.	-
TM4	The code is obtained.	-
TM5	The code can be used.	-

# 2.6 Evaluation and Recommendations

In the light of the results presented, all the PDKs tested confirm their mature shape, i.e., they all carry out the expected functionality correctly and they all run efficiently enough to work within the PIMCity platform, as reported in *Deliverable D1.4 Agile Report*.

In terms of resources used by each PDK, they all throw healthy results, exceeding by far the expectations of the tests. This shows that the software developed is not only production-ready, but also suitable for open-source environments where sometimes there is a lack of fine hardware and network bandwidth.

Given the response times obtained, the current PDKs can be used for several real-word use cases, including but not limited to integrating with Ad Servers and DMPs. Moreover, the platform seems highly scalable. However, if higher volumes are needed, horizontal scaling can be used as a technique to increase the availability and transactions throughput for each PDK. In such case, a load balancer is needed to receive the requests and forward them to different servers to be processed.

## 2.7 Conclusions on the demonstrator

The Continuous Testing Pilot showed to be a success. The design for the pilot allowed the Consortium to test each component while they were being developed, increasing their quality day by day. Tests could be performed without any major issue, and bugs were fixed right away.

All tested PDKs are fully functional, high quality and highly scalable if needed. Resources and environment configurations needed for any PDK to be run are not restrictive and are well available within the community.





Moreover, the documentation and OpenAPI specifications offered by each PDK allows the partners and external developers to integrate them easily with other PIMCity tools or even new software developed by anyone after the project has concluded.





# 3 B2B Testing

The second pilot aims at exploiting the technology developed during PIMCity in the B2B operations of ERMES. To this end, FW and ERMES have explored the possibility of involving business customers from their portfolios to test their interest in products combining preservation of security and privacy with training and increasing of users' awareness.

For this, the main goal of this pilot was to demonstrate the value of a subset of components in the B2B segment. ERMES has developed solutions for companies to help them protect their data while increasing their awareness with augmented information about the trustworthiness of web services employees' contact during their navigation.

## 3.1 Demonstrator

# 3.1.1 Demonstrator objectives

Differently from the previous pilot which aim is to demonstrate PDK components developed in technical work packages (WP2, WP3 and WP4), this demonstrator aimed at showing how ERMES integrated WP2's PPM and WP5's TT in its portfolio of tools for web protection, i.e., in a production environment. For this, we can consider this demonstrator reached TRL 8.

The final outcome of this pilot was two-fold: First, we evaluated the benefits introduced by TTs to an existing tool for web protection in terms of training and awareness. Second, we aimed at developing an exploitation plan for the commercialization of PIMCity's technology. This task, carried out by FW and ERMES, has already started and will continue after the end of the project.

## 3.1.2 TT Demonstration

The Transparency Tag is a tool similar to a Nutrition Label for food. The Nutrition Label provides the information about the ingredients, their provenance, intolerance risks, etc. of food. The Transparency Tag provides the privacy implications of visiting different web sites.

The TT has been designed and developed as an easy-to-understand interface to provide the user with information about the nature of the services contacted during the navigation. For each web service (e.g., a website, a mobile app, a data buyer), we have built a TT to present the information contained in the corresponding Privacy Metric, such as its owner, its purpose, the personal data it collects, etc. Apart from providing all the details, the TT also summarizes such information in scores - automatically computed by the analytics behind the Privacy Metrics - revealing the potential security and privacy risks associated to that service. TT also attempts to describe how transparent the service under exam is towards the user, i.e., whether crucial information such as data processing purposes, data controllers, contacts, etc. is provided publicly.

Although the TT design for B2C and B2B share some common points, the design for the business segment must respond to requirements which are different from those defined for private consumers. Indeed, for instance, TTs for private consumers need to address the problem of being easy to understand for user segments very different to each other, such as, e.g., tech-savvy, privacy experts, and digital illiterates. Differently, the TT design for B2B must consider very different requirements, such as the need to provide information about privacy and security postures of web services, including details regarding reputation,





categorization, presence in security lists, etc. Therefore, from a technical perspective, this demonstrator had to validate the design of TTs that ERMES has developed for its product (*Ermes for Enterprise*), and leveraging the set of clients involved.

## 3.1.3 Demonstration design

The demonstrator has been temporally divided in three periods: First, ERMES has implemented the B2B version of Transparency Tags in its main product, *Ermes for Enterprise*. Implementation has been followed by internal tests which helped ERMES to spot bugs and glitches in the presentation. Examples of TTs as integrated in *Ermes for Enterprise* are provided below. Similarly, as for EasyPIMS, TTs have been implemented as a webpage whose links are associated to tracking domains blocked by *Ermes for Enterprise* (more details on the design are available in deliverable D5.2).

OWNED BY Google - http://www.google.com		
TRACKER RANK PREVALENCE 8	SEEN ON THESE TYPES OF WEI	BSITES
	8.67%	Educational Institutions
	8.11%	International News
TRACKER REACH	7.73%	Streaming & Downloadable Video
<b>4.37%</b> of web traffic is tracked by YouTube	7.49%	Search Engines
2621 of the top 10 000 sites seen loading the YouTube tracker	5.18%	Educational Materials & Studies
	4.48%	Technology - Other
	3.25%	Marketing Services
OPERATES UNDER	2.73%	Online Shopping
<ul> <li>googlevideo.com</li> </ul>	2.69%	Advocacy Groups & Trade Associations
<ul> <li>youtube-nocookie.com</li> </ul>	2.69%	Uncategorized
voutube.com	2.07%	Government Sponsored
	1.79%	Magazines
• ytimg.com	1.7%	Banking
	1.56%	Video & Computer Games
TRACKING TYPE	1.32%	Philanthropic Organizations
local_storage () cookies ()	1.27%	Web Hosting, ISP & Telco
	1.23%	National News
TRACKING CATEGORY	1.08%	Personal Pages & Blogs
	1.08%	Entertainment News & Celebrity Sites
audio_video_player	1.04%	U.S. Government Resources
Enables websites to publish, distribute, and optimize video and audio content.	0.99%	Community Forums
	0.9%	Business - Other

Figure 1: Ermes for Enterprise Transparency Tags for www.youtube.com when this operates as a third-party tracker.





182.0 B of user data from tracker page         4.9 TRACKER         On average per page         98.41% PROPORTION OF TRAFFIC         Page loads from youtube.com on which tracking occurred         10 REQUESTS TO TRACKERS         Tracking request per page load	TRACKING METHODS LOCAL_STORAGE is a per-website key-value database contained in the browser which websites can use to store information to identify you. COCKES are files placed by the website, stored in the browser that is used to identify you to the website.
All trackers seen	Trackers Seen         Coogle         google.com         doubleclick.net         ytimg.com         googlevideo.com         youtube.com

Figure 2: Ermes for Enterprise Transparency Tag for www.youtube.com when this operates as a first-party website.

During the second phase of the demonstrator ERMES has run a beta testing campaign involving volunteering clients taken from its portfolio. In total, ERMES invited ten clients. Of these, six customers accepted to join the campaign and provided feedback which helped ERMES to further improve the implementation and evaluate the interest of clients for this kind of solution. More in details, ERMES has configured the TT as an independent module to activate on demand within its product. Hence, TTs have been enabled only for customers who accepted to join the campaign. ERMES then has chosen ten customers randomly, without considering size of the company, industry sector, or other parameters, and contacted them by phone to probe their potential interest in the TT solution. In general, we contacted the CISO, the IT manager or whoever oversaw the procuring and administrating security tools for the company<sup>2</sup>, asking them to organize a meeting to demonstrate TTs. To maximize the possibility for customers to accept to join the campaign, we specified that TTs was a new feature recently added to the product and activation was included in the existing subscription plan and free of any charge. At the end of this round of meetings, six clients accepted to activate TTs and to organize a second meeting after a few weeks to provide observations and feedback about TTs. All customers agreed to share their feedback, only if properly anonymized. The following table reports the anonymized list of companies along with some data.

Company ID	Industry Sector	Size	Revenue
1	Fashion	500-1000 (2021)	€ 111M (2020)
2	Chemicals	500-1000 (2021)	€ 248M (2020)
3	Energy Provisioning	10-50 (2021)	€ 59M (2020)
4	Manufacturer	100-500 (2021)	€ 48M (2020)

<sup>&</sup>lt;sup>2</sup> customers' contact information is available is ERMES' CRM system





5	Energy Provisioning	50-100 (2021)	NA
6	Banking and Financial Services	100-500 (2021)	NA

Table 1: ERMES' customers involved in TT beta testing campaign.

We resume the feedback we obtained from these early adopters in the following section.

As described in deliverable D1.3, the last phase of the demonstrator was meant to involve clients from FW's portfolio (tentatively about 50). ERMES and FW started working on the pilot in March 2022, but due to Fastweb's focus on its roll-out of 5G mobile and mobile-enabled fixed services on the target market segment in the same period, the pilot has been postponed. This shift in focus couldn't have been anticipated in the project planning phase. Interest in the pilot has not waned, though, and Fastweb and ERMES will conduct it as outlined in the exploitation plan (Deliverable 6.7). Furthermore, ERMES and FW are building the grounds for a broader collaboration which could lead to involve many more prospects than those planned. For this, at the time of writing, the results of this demonstrator have been obtained solely based on the feedback collected by ERMES during the beta testing.

# 3.1.4 Demonstrator results

Results of this demonstrator are based on the feedback that ERMES has collected during virtual and physical meetings organized with customers involved in the beta testing campaign described above. The number of involved customers is by no mean statistically representative, yet the results are meaningful from a qualitative perspective and substantially helped ERMES grasping customers' needs when it comes to create awareness and deliver technical information and intelligence.

We preferred to collect feedback in face-to-face interviews instead of using surveys to maximize the number of respondents. Moreover, face-to-face interviews allowed ERMES to drill down observations and comments. Indeed, surveys do not allow grasp details, feelings and nuances which are instead easy to capture in face-to-face interactions. Interviews have always involved two people from ERMES side. They took notes independently, and summarized them together at the end of the meeting to report feedback free from subjective biases.

We can divide the obtained feedback based on the four main questions we asked during the interviews: 1) what do you like of Transparency Tags? 2) What do you dislike? 3) What would you like to see improved/expanded? 4) How much would you pay for this service?

During these sessions, ERMES also received minor suggestions to improve the presentation (e.g., title alignments, colour choices, etc.) and to fix bugs. We omit to report these for the sake of brevity.

We summarize main feedback in the following.

## Q1: What do you like of Transparency Tags?

In general, all respondents welcomed TTs positively. They appreciated the information and details presented in the TT pages. All of them stated TTs allow to motivate why connections to trackers and other third-parties are blocked by *Ermes for Enterprise*, and let them quantify the actual amount of information collected by trackers and possibly leaked out of the





company. For three customers, data provide in TTs was very surprising as it shed lights on an ecosystem – trackers - they were not completely aware of.

Two of the customers stated they had never seen a web protection product providing such amount of detailed information, especially when it comes to privacy-related matters. More specifically, one customer appreciated this kind of intelligence since similar products typically provide detailed information only if attached to security alerts.

#### **Q2: What do you dislike of Transparency Tags?**

Although the general feedback was positive, all respondents only partially liked the presentation. They declared it can be improved, and data should be prioritized in a different way, although we registered some disagreement on what to emphasize. For instance, one respondent would prioritize technical information about tracking techniques. The same information was irrelevant for another customer. We conclude that TTs should be made customizable in a way that users can organize the submodules composing the TTs as they like.

The second main negative comment (four respondents) was that in its current form TTs provide information about trackers only. They expected to have similar information for other kinds of web threats for which *Ermes for Enterprise* provide protection, such as phishing, cryptomining and malvertising. Apart from creating awareness, this would ease security response and intelligence operations. This is interesting, but outside the scope of PIMCity.

The third main negative feedback (highlighted by three respondents) was that in its current form, Transparency Tags provide a snapshot of the blocked service which is not aware of the company scenario. Indeed, they do not report information about blocked services in connection with the user or the company. Customers stated they expected TTs to provide, for instance, information about personal data collected by the service specifically for the user involved in the communication, or for all users in the company.

Finally, two respondents stated that TTs represent a powerful tool, but that in its current implementation is useful to build awareness for *Ermes for Enterprise* administrators only. They would welcome a solution capable of transmitting the same information to the company employees as well.

#### Q3: What would you improve/expand?

All the feedback collected at Q2 requires to deeply revisit the TT design to be accommodated. However, we also collected suggestions which would help ERMES to improve TTs with little effort.

Two respondents noticed some information provided in TTs is not accurate or does not correspond to their observation. For instance, for some websites, the list of included trackers was not complete, as for others the list included trackers not actually connected to the website. Similarly, some trackers were reported to use local storage to store users' information, but apparently, they were not. This are technical information that keep evolving over time.

Other suggestions are related to augmenting the information currently provided in the TTs:





- Three respondents would like to integrate information about trackers with their Privacy Policies and/or other documents which should demonstrate GDPR-compliance.
- Two respondents would welcome more details on fingerprinting techniques used by trackers and the nature of data they collect.
- Two respondents asked to include tooltips to explain data.
- Four respondents asked to include scores summarizing the overall security risk associated to the trackers (as done in EasyPIMS).

#### **Q4: How much would you pay for TTs?**

Three respondents answered that, even if they appreciate TTs and believe in their potential, they will not pay extras to have TTs integrated in *Ermes for Enterprise*, especially in its current form. They would change opinion if TTs will be expanded to address other kinds of web threats or implemented as a tool to build privacy awareness for all company employees.

The other three respondents stated they would pay 5% to 15% more than what they currently pay for *Ermes for Enterprise* when the solution will be more mature (no bugs/glitches, better accuracy, and improved presentation).

## 3.1.5 Conclusions on the demonstrator

The final number of customers involved in this demonstrator was not enough to obtain any statistically solid conclusion. Nevertheless, the feedback we obtained was enough to confirm that Transparency Tags provide insightful information on the tracking ecosystems in the context on a system for web protection. If properly expanded, TTs have the potential to become a very useful tool to accelerate remediation and intelligence operations caused by web security incidents. Finally, ERMES is considering the idea of making them available to companies as tool to grow privacy awareness of their employees.





# 4 B2C Testing

The B2C pilot of PIMCity is the main pilot of the project. It has been designed to test all the components of the PDK integrated in the EasyPIMS platform with real users. In this section we describe the main actions performed to ensure the success of the pilot, and the results obtained after gathering the opinion of the final users.

In this pilot we demonstrate three different (and complementary) parts of the PIMCity technology:

The PDA: The Personal Data Avatar is a digital projection of the data stored by the user, and under full control of the user. In other words, the PDA is the interface between the user and the services. The demonstrator of the PDA aims at attracting a large number of users to the platform. Allowing them to upload their data and sharing it with third parties.

The Marketplace: The Marketplace allows third-party companies to interoperate with the data stored in EasyPIMS. Similar to a physical Marketplace, the EasyPIMS Marketplace facilitates the trading of user data (via the PDAs) and web services.

The Telco demonstrator: The telco demonstrator works separately from EasyPIMS. It allows to test how the integration with the huge amounts of data available (and continuously generated) at telco premises could be done by testing some of the most computational heavy components of the PDK. To this end it is divided in two smaller demonstrators: the privacy-preserving user profiling and the location history demonstrators.



Figure 3: High level design of the B2C demonstrator

Figure 3 shows the high-level design of the EasyPIMS platform and how it interacts with the different components to be demonstrated. The platform presents two main dashboards, each one of them corresponding with one of the components. On the one hand, the PDA is used by the final users to insert data, check information and complete tasks. All the data is saved in the Personal Data Safe, and the Personal Consent manager is used to specify their preferences. On the other hand, the Marketplace is used by data buyers to access to the data in the platform. It also first interacts with the Personal Consent manager, and only has access to data if the users consented to share it.





# 4.1 Demonstrator design, the PDA

The PDA is the way of the Internet users to interact with the EasyPIMS platform. In the last months of the PIMCity project we have designed a demonstrator that allows thousands of users to participate in the EasyPIMS platform.

We divided the demonstrator in three phases. The first phase, performed during the months of March to May 2022 was devoted to a closed beta version of the platform that was tested by the members of the consortium and 40 external members participating in focus groups. The second, and most important, phase was the opening of the platform to all users, and the starting of our strategy to attract and engage them using gamification and prizes. The final phase, corresponding with the last 15 days of August was used to obtain the evaluation of the platform from the users. In this phase, a survey was distributed to all the users registered in the platform. For a reference of the final implementation of the PDA we refer the reader to Deliverable 5.2.

**Focus groups**: Following the strategy designed during the whole project, the first evaluation of the PDA was done in focus groups with 40 external users that allowed us to improve the platform behaviour before it was open to the public.

The components of the Focus Group were recruited from among AUI associates. They volunteered to participate. Two virtual meetings were held with them. The first one to explain the objective of the project, the functionality of the platform, the system of prizes and sweepstakes to be developed. The second meeting was to present the different tasks and they were asked to finalize their duty in the Focus Group by completing a survey.

All of them have continued as beta testers during the next phase of the project.

During this phase at the beginning of June, the developers obtained direct feedback from external users that reported. That feedback allowed to fix several bugs in the platform (more than 15). Moreover, navigation patterns were changed to improve the usability

**Platform opening**: at the beginning of June 2022 the platform was publicly open to the public and advertised in different events and on Online Media Networks. The actions designed for the engagement of users are described in Section 4.1.1.

During this phase, an open communication channel was kept open with the final users through the email address <u>betatest@pimcity-h2020.eu</u> which is advertised on the platform itself. Some users reported different problems (and difficulties to use the website) that were quickly solved by the different developers. Some users used it to cancel their registrations (detailed figures provided later).

**Feedback gathering**: Finally, a survey was distributed to all the users in the platform. To ensure the participation of the users in the survey, a special 1900 points where added to those participating in the survey. With this number of points, the participation in the final draws (weekly, monthly and final) is assured if the survey is completed and the socio-demographic data is filled in.

331 users filled the survey at the end of August. Section 4.3.1 contains the results of this survey and the comments and suggestions reported by the beta testers.





# 4.1.1 User engagement

#### Public engagement and community building

One of the objectives of the project is to share experiences and results with the different stakeholders interested in issues related to personal data, privacy, and user-centric solutions such as those implemented in PIMCity.

The following table shows the different stakeholders that have been focused on during the execution of the project together with the actions developed during the project to achieve the participation and engagement of each one of them as described in D6.5.

#### End-user participation

End users have participated in focus groups, in testing prior to the launch of EasyPIMS and finally as beta testers

In order to reach them, we have worked with these groups:

- AUI associates we have contacted directly. This has allowed 160 associates to participate, 40 in the initial focus groups and 120 as EasyPIMS beta testers.
- Citizens from different countries that we have approached through a campaign of advertisements in social networks. They were given the opportunity to participate in the EasyPIMS platform as beta testers.
- Students from the universities in the project: POLITO, UC3M, KUL. Students have been reached through the different channels made available from each university.

As an incentive, 7,400 Euros have been distributed as prizes for weekly, monthly and a final sweepstakes as follows:

- During the first 14 weeks a 200 Euros Amazon gift card each week.
- During the months of June, August and September monthly sweepstakes for an iPhone 13.
- A final grand prize in consisting of a 2,500 Euros Amazon gift card.

The participation on the sweepstakes was awarded by the participation in different tasks, allowing the gamification of the platform.

As a result of these actions, a total of 3360 users have registered on the platform www.EasyPIMS.eu.

#### EasyPIMS tasks

Users registered in EasyPIMS website (<u>https://www.EasyPIMS.eu</u>) can access their personal dashboard. There, the welcome screen shows a list with the tasks they can perform at any time, the periodicity with which they can be activated, the points they generate and a button shown in green for those tasks I can perform at any time and in red for those I have already completed (the periodic ones are reactivated when the week or month changes).





To the right of the task panel the total accumulated points, the monthly total and the weekly total are shown and below there is a button that allows users to see the detailed history of all the tasks that have already been completed by each user.

Update your personal data	One time	50 Points	Task link	3080
Download the plugin	One time	200 Points	(Task link)	3300
Upload browsing data	Daily	10 Points	Task link	
Display 3 or more ads	Daily	10 Points	Task link	Points this month: 720
Set consents	One time	50 Points	Task link	Points this week: 410
Invite friends	Weekly	350 Points	Task link	
Upload location history	Monthly	200 Points	Task link	

Figure 4: List of tasks performed and to be performed by the EasyPIMS users

Following, we describe the different tasks to be performed.

#### Personal data task

For this task, users can enter personal data (socio-demographic and contact data) to share with third parties. To complete this task, users have to enter in the MyData section at least the information marked as mandatory (\*). This data can be updated at any time from the "MyData" menu option.

My Data

Here you can a	check and modify the data abou	t you:		
• Vou ogn opt	or your domograpich information whi	ab may be used by data buyers to filter au	dianaaa	
Check your I	browsing history that allows the system	m to build a profile with your interests.	ulences.	
<ul> <li>Import the log</li> </ul>	ocation history from your Google acco	unt so that data buyers could use it for the	ir need.	
The sharing     about the ar	of which information data buyers can mount of sites and services you brows	access, is controlled by the Consents you ed, the locations you visited, and the dem	give. In the bottom part of the page you can fir ograpich information you gave to the platform.	nd detail:
Contact informati	on	last name *	Gonzalez	
First name *	Roberto			
First name * Phone number	Roberto			
First name * Phone number Mandatory to partie	Roberto			







On the "My data" section the user can check and modify the data collected on EasyPIMS about her:

- The user can enter and change demographic information which may be used by data buyers to filter audiences.

- The user can check her browsing history that allows the system to build a profile with her interests.

- The user can Import the location history from you're her Google account so that data buyers could use it for their needs.

The sharing of which information data buyers can access is controlled by the Consents authorized by the user. In the bottom part of the page, the user can find details about the sites and services she browsed, the locations she visited, and the demographic information she shared with the platform.

The more data the user shares, the more likely it is that her profile will be attractive to these companies. Each time her data is submitted to a Data Buyer she earns points on her account to participate in sweepstakes.

#### Download Plugin Task

On this task, users are asked to install the plugin for Chrome. A plugin is an extension that is installed on the web browser to extend and increase its functionalities.



Figure 6: EasyPIMS plugin in the Chrome Store

The functionality of the EasyPIMS plugin (available for Google Chrome and Microsoft Edge) is to collect the browsing data to calculate users' interests and do it automatically by using the DKE module. This way, the platform can generate a profile of user's interests that the user has decided to share with companies participating in the platform.

This Plugin is specially designed for Google Chrome<sup>3</sup> (and it can be also installed in Microsoft Edge and Firefox). To download it, users have to click on the button "DOWNLOAD PLUGIN" and then click on "Add to Chrome". Completing this task adds 250 points to the account.

Once the plugin is installed, the browsing data collected by this plugin is displayed in "My data" section by clicking on the "Details" button of the "Browsing history". Here all the records collected are displayed sorted by date. The user can use this interface to delete the entries she does not want to share at any time.

## Display Ads Task

This is a daily task in which the platform shows the user three different ads and ask her to rate how much these match her preferences. The rating is done by marking from one to five stars each of the advertisements. Completing this task, awards the user with 10 points.

<sup>&</sup>lt;sup>3</sup> The Extension is available in the Chrome Store: <u>https://chrome.google.com/webstore/detail/EasyPIMS/mimpinmnbmpkiagdhljoillmjhhhapif</u>







Figure 7: Example of how the ads can be evaluated by the users

# Set consents task

On this task, users must configure the "My consents" section. Consents can be changed by the user at any time by activating the My Consents option from the main menu. In this section it is possible to select the preferences about which data to share for which purpose:

- Preferences are defined by generic purposes and corresponding consents. Based on these, the user authorizes to share her data on the marketplace where data buyers can buy it.

- Interest contents allow users to specify their interests from the list that advertisers will use to show personalized ads.





My Consents



Figure 8: My Consents section of EasyPIMS

The histograms show an overview of the personal interests as computed from the navigation history.

Private companies, public organizations and educational institutions may buy data for specific purposes. When a user ticks a box, she gives permission to share her data for the specific selected purpose. For each organization requesting such data, the user receives some points.

#### Invite friends task

In this task, users should invite their friends to participate in the platform by sharing a personalized link (aka, *referral* link).

For each friend registered following a personalized link that completes more than one tasks, the inviting user receives a total of 350 points. To accumulate these points, it is necessary that the friend, in addition to registering, earns at least 100 points in her account. The maximum number of friends per month is limited to three (after the third friend, no more points are earned).

#### Upload location history task

This extra task allows users to share the geolocation data generated when using Google Maps. To upload this data to EasyPIMS, they must first download it from Google Maps and then upload it to EasyPIMS platform.

In this video we show the complete process of how to do it: https://youtu.be/\_6OBZrHJ\_vM

Completing this task adds 250 extra points to the account.





The following table summarizes the different tasks available to the users.

EasyPIMS Tasks	Description	Point s	Repeata ble	Frequen cy
MyData: Update your personal data (Contact and	Enter contact and socio-			
Sociodemographic data)	demographic data	50	No	Once
	Select your preferences about which data to share for which			
Set consents	purpose	50	No	Once
	Invite friends using your referral link			
Invite Friend	and complete the weekly tasks	350	Yes	Monthly
	Download and install EasyPIMS plugin in the browser			
Plugin Installation		250	No	Once
Location History	Upload the location data from Google data	200	Yes	Monthly
Daily Browsing	Browse with the browser installing the plugin	10	Yes	Daily
	This is a daily task in which we show you three different ads			
Display and evaluate suggested ads	and ask the user to rate your interest in each ad shown.	10	Yes	Daily
Evaluation Survey	Complete project evaluation form	1900	No	Once
	For each organization that requests a user's data, the user			
Marketplace offers	will receive points.	50	Yes	None

#### Advertising

In order to invite a larger mass of users, we designed an advertisement campaign on different social networks. The campaign aimed at advertising EasyPIMS has been carried out on Facebook, Instagram, and LinkedIn.

The ads that have been published are shown below. Users were invited to visit the platform website and to register on it. A total of 12000 euros was invested in the campaign. This campaign has been active from June to July 2022





English Ad	Spanish Ad	Italian Ad
Do you want to better understand how to value your data online? Participates in the European	¿ Quieres dar valor a tus datos en Internet ? Participa en el	Vuoi capire meglio come valorizzare i tuoi dati online?
Chapter to wid 3 <i>Uncert</i> Id pift cards of 2000 and 1 gift card of 2.5000 www.estypimt.eu Toto and the state of 2000 www.estypimt.eu Toto and the state of 2000 wwww.estypimt.eu toto and the state of 2000 www.estypimt.eu toto and the state	proyecto Europeo EasyPIMS y gana premios www.easypims.eu Europeo Europ	evinga i Simbane II. 14 carte regalo de 2006 e 1 carte regalo de 2.5002 www.eatypine.eu

Figure 9: Campaign Ads in three different languages

## Sweepstakes

To incentivize the participation of users as beta testers and contributors of their personal data in EasyPIMS, a reward system has been designed with a total of 7,400 Euros, which are distributed in different sweepstakes that can be accessed if a previously established number of points are obtained by performing the tasks described in the previous section.

EasyPIMS		USERS AREA	FAQS SWEEPSTAKES INFO English	~
Next Sweepstake	es			
SWEEPSTAKE	PRIZE	LOTERY DATE	PARTICIPANTS	
Weekly (22/08/22 - 28/08/22)	200€ Gift Card	01/09/2022	328 Participants (list)	
Weekly (29/08/22 - 04/09/22)	200€ Gift Card	08/09/2022	Participants list 06/09	
Monthly (August)	Iphone 13	08/09/2022	Participants list 06/09	
Final sweepstake	2500€ Gift Card	15/09/2022	Participants list 13/09	
Sweepstakes wir If your User ID is among the SWEEPSTAKE	INERS winners write to us at prizes@easypin PRIZE LOTERY DATI	<mark>ns.eu (including your User ID, F</mark> i E PARTICIPANTS	irst and Last name) WINNER	
Week (15/08/22 – 21/08/22) First prize 79987 National Lotery Website [English] [Spanish]	25/August/2022	<u>305 Participants (list)</u>	Winner EasyPIMS User ID: c0bc3152-75f7-4267-9087-5e60588539b4 Number 072 on the list Calcul: 072 = Module (987/305))	
Week (01/08/22 – 07/08/22) First prize 79073 National Lotery Website [English] [Spanish]	11/August/2022	<u>132 Participants (list)</u>	Winner EasyPIMS User ID: 9314648f-5d41-49de-b58a-867c73f6024d Number 073 on the list	
Month July First prize 49221 National Lotery Website [English] [Spanish]	04/August/2022	410 Participants (list)	Winner EasyPIMS User ID: e19f614d-cb85-4211-9263-7ff39b8b9aa6 Number 221 on the list	







Each time some of the tasks proposed in the EasyPIMS platform is completed, a given number of points is assigned. Points are accumulated for the sweepstakes described in the following. The table below shows the allocation of points for each task and the frequency with which each task can be repeated to accumulate points.

A weekly sweepstakes is scheduled for each of 14 weeks of the demonstrator. Each of them endowed with a 200 Euro Amazon gift card, in which those users who have accumulated at least 100 points in the previous week participate.

Three monthly sweepstakes, each of them endowed with an iPhone 13, among all users who have completed at least one task in the previous month and a final sweepstakes among all those who have achieved a total of 2,000 points since the beginning of the experience.



# Figure 11: Prizes awarded by the sweepstakes

The webpage https://www.EasyPIMS.com/sweepstakes reports all the competitions, both the future and the completed ones, together with their winners. For the sake of transparent, the page reports the codes of winning users, and the list of users who have achieved the objective to access the competition. The page also describes how the winner is chosen: to ensure that the choice of the winners is absolutely made randomly and cannot be manipulated, the winner is chosen from the winning number in the Thursday competition carried out by the Spanish National Lottery, taking the last digits of extracted number, the winning number is chosen from the list of user IDs of those users who have achieved the necessary points for that draw.

The list of participants is made public at least two days before the day of the competition, so that the prize cannot be assigned to a user who is not on the list. Only the user ID assigned by EasyPIMS is published, thus we do not publish any personal data of the participants.

This system guarantees that the choice is fair and that the winner is designated by the Spanish State Lottery administration, which is an organization external and independent from the PIMCity project.

## Other communication actions

Communication actions to publicize EasyPIMS have focused on the following activities:





- Posting on social networks (Facebook, Twitter and LinkedIn) about the project: posts have been published both of the announcements and of the participations in the different events where the platform has been presented. These posts have been replicated in social networks pages of the project partners.
- Public presentation of EasyPIMS platform at international events: CPDP 2022 in Brussels, MyData 2022 in Helsinki, Interact 2022 in Madrid. Organization of a business event to present the platform to companies in the data sector through IAB.
- Publication of a newsletter specifically about EasyPIMS and publication of an entry in the BDVA and MyData Global newsletters that reached thousands of project professionals.
- Publication of specialized papers and articles.
- Elaboration of 2 press releases that have been sent to international media and news agencies.

The details of all these actions are provided in deliverable D6.6.

# 4.1.2 Demonstrator Evaluation

During the development of this demonstrator 3360 users registered in the EasyPIMS platform. From there, 46 users requested the deletion of their account via the functionality offered by the platform itself. Figure 12 shows the evolution on the number of users registered in the platform. As shown, the initial testing started in mid-May, the opening to the Focus Group in the 20<sup>th</sup> of May, and the grand opening in June 1<sup>st</sup> with the start of the advertisement campaign on Facebook which brought the largest number of users.



Figure 12: Evolution of total number of registered users in the platform

While more than 3600 users registered to the platform, only 910 shared some data, of which 825 entered their country of residence.







Figure 13: Types of data shared by user

Figure 13 details which type of data such users inserted in the platform. As it can be observed, of the 3600 users, only 910 inserted their personal data at the end of the experimental period, 337 downloaded and installed the plugin that registers the browsing history, and only 205 completed the upload of their location history on the platform. These are the numbers as observed at the end of the period, after several reminders were sent to all registered users (three emails were sent to all users at different periods of time, namely on July 22<sup>nd</sup>, August 16<sup>th</sup>, August 27<sup>th</sup>).

Figure 14 details the number of users per country, while Figure 15 shows the growth of those numbers over time. Considering the countries, Spain and Italy result the two countries where most users registered from. This clearly reflects the bias in the consortium whose partners pushed EasyPIMS though their network of contacts. Interestingly, South America countries are also quite well represented. This is due to the value of the prizes that are likely very high in those countries. Indeed, the effectiveness of the Facebook campaign was particularly high in those countries, as clearly visible in Figure 12: Starting from the 1<sup>st</sup> of June, the growth of users in those countries is quite sizeable also when compared to the main language for these South American countries is Spanish, one of the languages which we used the most to advertise EasyPIMS on social networks.

In June 15<sup>th</sup>, we suspended the ad campaign on Facebook, and this caused the growth to stop in the majority of the countries. In June 15<sup>th</sup>, POLITO sent an email to all the students (about 35,000 emails) asking them to register to the platform. This caused the sudden growth of the number of new users (that also entered some data) in Italy (and foreign countries for not Italian students – see the growth in the other countries curve). Other major growths are also visible in July 2022 in Spain at the time when AUI and IAB Spain solicited some users to join the platform.

Unfortunately, we observe almost no impact in the correspondence of those solicitation emails that were sent to all participants on July 22<sup>nd</sup>, August 16<sup>th</sup>, August 27<sup>th</sup>. Notice that some of those emails might be tagged as spam by mail servers.





Figure 14: Number of users providing country information



Figure 15: Growth in the number of users over time for the most active countries for those users that entered such information

This testifies that, despite our effort in on boarding users and in designing a simple and straight forward user interface, the interest of users toward the PIMS seems to be limited. Considering for example the POLITO students, we can estimate that only about 100 students out of about 36,000 emails (2.8%) joined the platform. Considering the cost of the advertisement campaign on Facebook and the acquisition of users that this generated, we can estimate a cost higher than 3.33 Euro per user (12,000/3600), which grows to at least 13,18 Euro (12,000/910) per users considering those that entered some information in the platform. Notice that these figures are a lower bound since we cannot distinguish the channel that users adopted to join the platform, and that email sent to students were free of charge for POLITO.





## Targeted advertising

The users were also asked to check and evaluate ads served using the profiles calculated by the EasyPIMS platform for the different users. In order to serve a targeted ad to the users we need to know their profile. And to calculate their profile we need the browsing information provided by the Chrome Extension.

Installing the Chrome extension was one of the tasks providing more points (200 points for the installation plus 10 for each active day). It made 1750 users (half of the users registered in the platform) to visit the installation webpage. However, only 348 eventually installed the plugin. We conjecture the installation of the plugin is complex for the average user. It makes more difficult to complete the task. Interestingly, this downside would be solved in a commercial implementation of EasyPIMS into a telco provider, since they would not require the users to install anything.



Figure 16: Number of daily active users of the Chrome Extension

Figure 16 shows the number of daily active users of the plugin since its first publication. As the number of users grows in June, it also starts growing from the initial 25 active users obtained after the focus groups. We observe a peak of close to 200 active users per day at the beginning of July that slowly decrease during August (probably, due to the holidays season) to slightly recover at the end of August.







Figure 17: Number of ads served by day

Each of the users can visits the EasyPIMS website daily to observe and evaluate up to three ads. Figure 17 shows the number of ads served each day during the three months of the experiment. We observe a peak close to 300 ads (that is, 100 users observing ads) at the beginning of June. Moreover, we observe a clear weekly pattern, with users participating much less during the weekends.



Figure 18: Average ad evaluation for different experiments.

Finally, Figure 18 shows the evaluation provided by the users to the ads they observe. We designed 2 different experiments. In a first phase, the three adds shown to the users were generated using their browsing navigation in the previous 20 mins. Those, the advertising should be very related to users' current interests. Then, in a second phase, we decided to show users three ads corresponding to three different profiles: The profile dynamically generated with the last hour of navigation, the profile dynamically generated with last day of navigation, and a profile, previously generated with the navigation patterns of the previous day.

As shown in the figure, we cannot observe great differences between the experiments and the evaluation of the users remains in average a little over 3 stars. While the results are





acceptable, we observe most of the valuations were either 1 or 5 stars. Then, we conjecture most users probably do not really evaluate how good the ads were for them but directly chose an option to complete the task.

# 4.1.3 Users feedback

The evaluation here collected has been carried out through an online survey that has been completed by a total of 331 users and that has been available in English, Italian and Spanish.

The first three questions are general and then the user was asked to rate each task to be performed in terms of comprehension, usability, security or trust and finally was given the option to make a free comment on each of them.

The survey questions and results are listed in the following table:













5.1) Rate from 1 to 5 if you understood this task ("**Download the plugin**": 1 if you did not understand anything and 5 if it was easy to understand)

5.2) Rate from 1 to 5 how easy it was to complete this task ("Download the plugin": 1 if it was very difficult and 5 if it was very easy).

5.3) Rate from 1 to 5 how safe you felt performing this task ("Download the plugin": 1 if it was not very safe and 5 if it seemed very safe)



5.4) If this task ("**Download the plugin**") has been a problem for you or you have any comments regarding this task, please let us know:

- Since I installed the plugin, popups often appear requesting login. Since they appear several times a day interrupting browsing, they are quite annoying.

- Doesn't work on mobile

- I use a browser other than Chrome

- The information provided when viewing the details is not very useful for the user

6.1) Rate from 1 to 5 if you understood this task ("**View 3 or more ads**": 1 if you did not understand anything and 5 if it was easy to understand)

6.2) Rate from 1 to 5 how easy it was to complete this task ("View 3 or more ads": 1 if it was very difficult and 5 if it was very easy).

6.3) Rate from 1 to 5 how safe you felt doing this task ("View 3 or more ads": 1 if you felt it was not very safe and 5 if you felt it was very safe)



6.4) If this task ("View 3 or more ads") has been a problem for you or you have any comments regarding this task, please let us know:

#### - The advertisements are continuously repeated











8.1) Rate from 1 to 5 if you understood this task ("**Invite friends":** 1 if you did not understand anything and 5 if it was easy to understand)

8.2) Rate from 1 to 5 how easy it was to complete this task ("Invite friends": 1 if it was very difficult and 5 if it was very easy).

8.3) Rate from 1 to 5 how safe you felt doing this task ("Invite friends": 1 if you felt it was not very safe and 5 if you felt it was very safe)



8.4) If this task ("**Invite friends**") has been a problem for you or you have any comments regarding this task, please let us know:

- There should be no monthly limit

- I never share what I do

- When my friend filled out the form from the link I forwarded to him, it was not clear if he had actually received it from me or if he was filling it out as a new ordinary user

- I did not use this task.

9.1) Rate from 1 to 5 if you understood this task ("**Upload location history**": 1 if you did not understand anything and 5 if it was easy to understand)

9.2) Rate from 1 to 5 how easy it was to complete this task ("Upload location history": 1 if it was very difficult and 5 if it was very easy).

9.3) Rate from 1 to 5 how safe you felt doing this task ("Upload location history": 1 if you felt it was not very safe and 5 if you felt it was very safe)



9.4) If this task ("**Upload location history**") has been a problem for you or you have any comments regarding this task, please let us know:





- Complicated to go to google to download data. This should be automatic as well.

- I did not subscribe to this option

- I followed the directions and the successful message came through, but the points never loaded

- Did not work

- It is not possible to complete the task.

- The task cannot be completed.

- Chrome would not let me download the locations. I had to download them with another browser

- I struggled to download the locations from Chrome. I had to use a different browser

- I followed the instructions and the words "loading performed" appeared, however the locations never loaded

- I did not perform this task because I tend to disable geolocation so as to limit the amount of data collected and transferred to Google (both to save battery power and because the latter is quite invasive).

10.1) Rate from 1 to 5 if you understood this ("Transparency Tags": 1 if you did not understand anything and 5 if it was easy to understand)

10.2) Rate from 1 to 5 how useful are "Transparency Tags" for you (1 if you felt it was not very util and 5 if you felt it was very util)



10.3) If this tool ("Transparency Tags") has been a problem for you or you have any comments regarding this task, please let us know:

Most of the websites that are displayed do not have any score. I don't know what they are for.

- The system is supposed to calculate automatically the webs that are shown do not have them and some of them are very well known sites.

## - They don't seem to work

11) If you have any other suggestions or comments, please let us know:





- Making it easier and more transparent

- Using nick instead of id in the winners list would be better

-Advertising and position uploading should work better; "fake" ads are unrealistic

-I noticed that despite that we had given consents on the use of preferences, the advertisements were not how many specifically related to them.
- Some actions such as the loading of geolocation data are very complicated and should be automated.

- The detailed information given to the user is not very useful and should be grouped by domains and filtered.

 The user should be informed about the Data Buyers to whom his data has been delivered.

The proposed tasks have been, in general, well understood by users and easy to use (95%) with the exception of the loading of geolocation data.

In terms of trustworthiness, the rating drops somewhat as only 67% consider the experience to be trustworthy or very trustworthy.

In the specific tasks of viewing advertisements and loading geolocation data is the one that has generated the most negative comments due to its non-functioning in some cases, to the fact that the advertisements shown are not of interest or because the process is complex.

The detailed information provided to the user on browsing history, geolocation data, etc. should be improved to make it more useful.

Technical improvements have been pointed out as being able to use the application with other browsers and mobile devices.

In terms of transparency, the user should be informed of the Data Buyers to whom their data is given.

## 4.2 Demonstrator design, the Marketplace

The Data Marketplace is the way the Data Buyers interact with the EasyPIMS platform. In the last months of the PIMCity project we have designed a demonstrator that allows multiple companies participate in the EasyPIMS platform and acquire data from users.

Also for this demonstrator, we divided the demonstrator in three phases. The first phase, performed since March to June 2022 was devoted to a closed beta version of the platform that was tested by the members of the consortium. The second, and most important, phase was the opening of the platform to all the buyers, after having on boarded several end users in the PDA to be able to build interesting audiences. The final phase, corresponding with the last 15 days of August was used to obtain the evaluation of the platform from the buyers brought by IAB. In this phase, IAB asked for direct feedback to those users registered in the





platform. For a reference of the final implementation of the Marketplace we refer the reader to Deliverable 5.2.

**Beta Testing:** Following the strategy designed during the whole project, the first evaluation of the Data Marketplace was done by the project partners, gathering all the feedback about bugs and UX suggestions that allowed the developers to improve the platform behaviour and usability before it was open to the public.

**Platform opening:** At the beginning of June 2022 the platform was open to the public and advertised in different events and workshops. The actions designed for the involvement of data buyers are described in the following section.

During this phase, a communication channel was kept open with the data buyers through an email address. Several buyers reported different problems when buying data that were quickly solved by the developers.

**Feedback gathering:** Finally, feedback was gathered from eight different data buyers that participated in the platform apart from the project members themselves. More details are available in the section Data Buyers Feedback.

## 4.2.1 Data Buyers involvement

IAB Spain has been the most active partner in this sense as it was one of its main roles during PIMCity Project. The strategy for the involvement of data buyers has been based on the following points:

- Provide data buyers with information about the project and the operative of EasyPIMS platform and PIMCity technology and modules through newsletters or general mailing lists.
- 2. Organise data buyer's workshops or events to present the project, its preliminary results and demonstrate the functioning of EasyPIMS platform.
- 3. Organise, with interested data buyers, one-to-one live demos for getting a closer look to EasyPIMS platform. Select the most interested data buyers for giving them access to EasyPIMS marketplace for testing purposes.

**Providing information to data buyers.** In this regard IAB Spain has a database of more than 300 hundred companies and 9000 individuals interested in buying end-users' data. Data Buyers had been constantly informed of the development and benefits of the project. 18 newsletters with information regarding PIMCity had been communicated to this audience. This has also a strong focus in getting feedback through simple surveys and communicating activities of PIMCity project for getting a better involvement of data buyers.



Figure 19: Examples of pieces of information provided to Data Buyers

**Workshop Organization.** In this regard IAB Spain has organised 4 workshops to promote EasyPIMS platform and demonstrate its results. The pandemic situation has forced conducting the first 3 Workshops via online platforms. They were communicated through IAB Spain newsletter and social media accounts, but also through IAB Europe newsletters as well as the official communication channels of the project. Only Data buyers were impacted through these workshops. In total more than 500 data buyers had been lively informed through the following workshops.

<u>PIMs Business Opportunities- 23 February 2021.</u> The goal of the first workshop was raising awareness about key elements of PIMS and how they are intended to be promoted through the European Union. During the workshop, data buyers had a first glance of how user-centric personal data management models could help their operations. The session also had a strong legal and ethical perspective analyzing how new and future data regulations and new models of personal data management will be going to be promoted in the EU. UE DG Connect members (Directorate-General for Communications Networks, Content and Technology) and the EDPS (European Data Protection Supervisor) joined the workshop to represent the public sector. IAB Spain, NEC, Onecub, Telefónica, Vastuu Group, Wibson were the speakers of the workshop. 93 companies attended this workshop having also more than 15000 impressions through IAB Spain social networks and more than 1000 visualisations through LinkedIn.







Figure 20: Workshop for "PIMs Business Opportunities"

<u>ZERO PARTY DATA. How to build a PIMs System and how to Benefit from EU PIMCity Project- 21 July 2021.</u> The goal of the second Workshop was highlighting key elements of PIMS and how they can help data buyers and advertisers in its daily operations. An explanation of what was the goal of PIMCity was given as well as an explanation of the overall operation of EasyPIMS and the different development kits that the project was developing. IAB Spain, Politecnico di Torino, Telefónica, Universidad Carlos III de Madrid were the speakers of the same. 27 companies attended this workshop having also more than 6000 impressions through IAB Spain social networks. The Workshop was also published on YouTube having more than 70 post event visualisations.



Figure 21: Workshop "Zero Party Data"

 <u>IAB Spain Digital Regulation Congress- 2 February 2022.</u> This is one of the biggest Spanish events on the digital regulation landscape, is promoted by IAB Spain and PIMCity project was its main sponsor. PIMCity project was presented by its technical coordinator Roberto González (NEC) that gave an overview of the project and an explanation of EasyPIMS operative. From a public perspective the Spanish data





protection authority (AEPD) through its director and the ministry of economic affairs were speakers of this event as well as private corporations such as Google or Onetrust. More than 500 individuals where registered in the event and more than 300 actually attended it, being 220 attendees representing the Data Buyer ecosystem. The event was published and communicated by Spanish publishers with interest in marketing and privacy and had an estimated audience of over 100,000 individuals reached through online newspapers and more than 5 million impressions in social networks.



Figure 22: Roberto González (Technical coordinator of PIMCITY) presenting at the IAB Spain Digital Regulation Congress

<u>INTERACT 2022- 26 May 2022.</u> This is one of the biggest European events concerning digital advertising. The event was hosted in Madrid and organised by IAB Europe and IAB Spain. PIMCity and EasyPIMS had a key role in the event as its technical coordinator Roberto González perform a keynote speech promoting the project and platform and inviting data buyers to test the platform and adopt similar PIMs models based on the work done in PIMCity project and in particular on its technical components. The event had more than 275 stakeholders attending the same and was recorded for streaming purposes.







Figure 23: Roberto González (Technical coordinator of PIMCITY) presenting at the IAB Interact 2022

#### Trials and testing.

More than 20 private corporations asked for more information or contacted to IAB Spain to know more about PIMCity. These include well-known search engine companies, financial services, insurance companies, AdTech industry and other companies developing business models with a PIM-oriented philosophy. IAB Spain had private sessions showing the insides of EasyPIMS platform, solving doubts and asking for feedback. A follow up was done after the sessions and some data buyers where keen on testing EasyPIMS marketplace platform.

## 4.2.2 Demonstrator evaluation

During the development of this demonstrator **16 data buyers** registered in the EasyPIMS platform.

Such data buyers executed a total of **81 transactions** in which they could ask for many data types in the same operation (like contact information, socio-demographics, location history, etc.).

Counting (non-exclusive) transactions per data type, we find:

- Contact Information: 13
- Socio-demographics: 41
- Location history: 10
- Interests: 25
- Browsing History: 5

A total of **1243 end users** were reached by these offers. Each end user participated in an average of **5 transactions** and got awarded with **50 points** per transaction. A total of **350,400 points** were granted for these data offers.

## 4.2.3 Data buyers' feedback





IAB Spain has carried out several actions to gather feedback from data buyers. IAB Spain has approximately 200 associates in the data buyer category, with approximately half of them being international companies operating throughout Europe.

One of the actions that was carried out in order to prioritize the technical development of EasyPIMS components was conducting a survey for data buyers asking which type of user data they were most interested in. The survey asked about four different types of data:

- Browsing history
- Demographic data and self-reported user interests
- Financial or banking data
- Social Media data

The survey was distributed over a two-month period to IAB Spain members, corresponding to the data buyer side, both in specific working groups and through newsletter communications. The survey had 67 responses with respondent's interests divided into data categories as depicted in Figure 24:



# Figure 24: Data Buyer's preferences when getting data directly from users 22 data buyers were most interested in user demographic data and self-reported interests

Based on the hypothesis that financial data and social media data are more difficult to obtain and contain a wealth of information, we thought these could be the most relevant categories for data buyers. Instead, the results showed a greater interest in demographic data as well as in user navigation. Even if not robust from the statistical perspective, this survey helped PIMCity partners to prioritize the technical development of the offers related to the most interesting information for data buyers.





During both the data buyer's one-to-one tests and the final test of EasyPIMS Marketplace, IAB Spain has gathered feedback from its members, being the most relevant information the following:

Feedback obtained through one-to- one beta testing	Feedback obtained through final EasyPIMS Marketplace
Difficulty in using data other than for research purposes or for direct email contact with users.	When placing an offer for a dataset and then placing a second offer in EasyPIMS Marketplace, the results are not negativized, so there is a possibility that the second offer returns data already acquired; therefore, it is not the most efficient system for data buyer's budget.
Transparency Tags and its measurement through stars does not reflect all the security measures and data utilization purposes of data buyers so the score may not accurately reflect their operations which could damage its brand image.	User consents cannot be downloadable once data is obtained from an offer in the Marketplace. Currently the consents are only available on the platform with no option for data buyer's visualisation. In any case, the system facilitates compliance as data can only technically be obtained once the user has parameterized their consents.
EasyPIMS ad system is not 100% compatible with the current use of data for online advertising.	Marketplace Dashboard containing little information. Regular advertising platforms provide data buyers with different ways to visualize the dashboard, such as graphs based on the different types of data acquired, data efficiency and more.
Low expectations about users engaging in the long term in this type of platform, as it requires a certain amount of commitment. This also leads to a lower volume of users and therefore less interest in data buyers.	The data that can be obtained through EasyPIMS is not linked to the user id, which would be useful in order to being able to connect different offers or types of data.
Low understanding of how the data valuation system operates, as there is no great difference in price between buying one type of data over another.	It would be convenient for a better Marketplace interface to be able to put names to data offers





JSON format when downloading data is recommended due to its standardization in data buyers.	It would be convenient to be able to sort and filter directly on the platform the obtained datasets.
Consent to "research activities" or to "commercial purposes" might be too generic for some data processing activities.	It would be convenient to be able to visualize the obtained datasets directly in the platform as well as when downloading them through the JSON file.
When completed the data offer it is desirable to have the option for checking which datasets you have downloaded previously, as the download is asynchronous when datasets are big enough.	When placing offers the platform filters the users but data buyers cannot know the total number of users available.
Being able to connect different accounts belonging to the same data buyer will be desirable.	There are times when bids get stuck and are not completed.
	When filtering the audience by job in a data offer, sometimes it returns data and other times it does not.
	Deleting some data offers from EasyPIMS data buyer's Marketplace might be desirable due to system failures or others.
	Sometimes the offers in the Marketplace are completed, but do not return any data, in any case the credits are not consumed.

Some of the feedback from data buyers has served PIMCity to improve the platform and therefore making it more useful when placing offers for data. Also, the information provided from data buyer's side have led PIMCity to consider different ways of solving problems, not finding remedies in some cases given the current state of technology, the actual functioning of online advertising and other data buyers and the lack of user's interest. The problems related to the platform are the ones PIMCity tried to work on, solving issues in the creation of offers or when obtaining data and others, however, data buyers are accustomed to deal with advertising platforms that could potentially reach all the audience over the internet and are already integrated with the main advertising technical protocols; That leads to problems like the lack of expectations in the vast majority of data buyers, as they follow other business





models, or the difficulties integrating with ad spaces and ad platforms that could not be solved within PIMCity technology.

## 4.3 Telco demonstrator

The final demonstrators have been designed to show how the PIMCity technology can be integrated directly with the data already present at the Telco operators. While socio-demographic data can be straightforward to obtain and transfer by the network operators, other data can be challenging given its size and dynamicity.

In this demonstrator we focus in two kinds of data especially valuable for the Telco operators: Users browsing history and user location.

The user browsing history can be captured by the telcos in very different ways. The easiest way is to collect the DNS queries performed by the users (that typically rely on the operator DNS server). However, other complex options, such as eavesdropping the user connection could also be possible if the permission from the user is granted.

In the case of the User location, the telco operators can triangulate the users to know their real position, or at least can know the antenna to which the users are connected.

# 4.3.1 Browsing history demonstrator: Privacy preserving user profiling with Telco data

In this demonstrator, we show the operation of the Personal Privacy-Preserving Analytics (PPPA) and Data-Knowledge Extraction (DKE) modules in a real use case. Using a dataset provided by the Fastweb partner, we showcase a complete processing pipeline, by means of which we anonymize a stream of data coming from a Telco operator, and use the anonymized data to build privacy-preserving user profiles. Our experiments show that the cooperation of the PPPA and DKE modules allows a data curator to extract meaningful and useful user profiles from data, while at the same time preserving the privacy of the involved users/data subjects. The demonstrator builds on the following steps.

**Data Collection.** This demonstrator employs data collected in a production network of a Telco operator. The partner Fastweb collected a dataset in the form of a trace of DNS queries captured from a subset of their customers. Using the logging capabilities of the DNS server software, Fastweb captured a 6-minutes long trace containing all the DNS resolutions issued by a population of their customers. Given its nature, this DNS trace indicates which websites are accessed by the monitored data subjects and the time of its access. While this information is valuable for various goals (i.e., user profiling), it contains sensitive information, as the list of websites visited by users can disclose personal characteristics (e.g., political or religious feelings). To this end, we took care of storing this dataset within the Fastweb's premises using a virtual private data center whose physical access is allowed only to Fastweb staff and remote access restricted only to the strictly necessary personnel of the consortium. The main characteristics of the dataset are summarized as follows:

- Size of the DNS trace: 4 GB
- Number of queries: 24 million
- **Duration:** 6.5 minutes
- Involved users/data subjects: 458,000





- Median queries per client: 20, IQR: 6-59
- Unique Domains: 2.2M domains, 716k second-level domains

**Data Anonymization:** we anonymize the dataset using the PDK module called Privacy-Preserving Analytics (PPPA). It includes various algorithms for data anonymization. The input data is a set of records, in the form of *(time, client-IP, domain)*, each representing a user identified by her *Client-IP* accessing a *domain* at a certain *time*. The domain represents the visited websites, as extracted by truncating the *path* part of a *URL* or as extracted from *DNS* traffic. We do not need to access other users' data, e.g., content of *cookies*, the full visited *URL* or personal details (age, gender, etc.). The anonymization operates on a twofold basis, anonymising:

Client-IP: the client IP address is the most privacy-sensitive fields, as it allows to identify the subscriber and link the records to the person generating it. To anonymize the client IP address, we will resort on the state-of-the art technique, called the CryptoPAN algorithm<sup>4</sup>. CryptoPAN replaces IP addresses with pseudo-encrypted copies while maintaining the network prefixes. In other words, the client IP address will be substituted by another, pseudo-randomly chosen IP address. CryptoPAN will ensure us that two IP addresses belonging to the same subnet will still belong to the same subnet in the encrypted space. Without knowing the encryption key, it is impossible to unveil the original IP address given the encrypted one. CryptoPAN encryption keys can be static or randomly rotated at fixed time intervals. In the following image, we provide an example of original and anonymized IP addresses.

IP address	Anonymized IP address
10.1.3.143	117.14.240.136
10.1.6.18	117.14.246.18
192.168.1.2	252.103.242.113
212.204.214.114	228.71.168.109
212.204.214.114	228.71.168.109
10.1.3.143	117.14.240.136
10.1.6.18	117.14.246.18
10.1.6.18	117.14.246.18
192.168.1.2	252.103.242.2
192.168.1.1	252.103.242.1

# Figure 25: Example of anonymized IP addresses

• **Domain**: while domains are not directly user identifiers, they can threat users' privacy if not handled correctly. Indeed, infrequent domains can be uniquely traced back to limited sets of users, thus jeopardizing the anonymization of users' identifiers (i.e., their IP address). In the literature, it is widely known that removing the user's identifiers (name, social security number, phone number, etc.) is not sufficient to make a dataset anonymous. Indeed, an attacker can link a user's apparently harmless attributes (the list of visited domains in our case) to some background knowledge. In this way, the attacker can re-identify the person and gain

<sup>&</sup>lt;sup>4</sup> Fan, J., Xu, J., Ammar, M. H., & Moon, S. B. (2004). Prefix-preserving IP address anonymization: Measurement-based security evaluation and a new cryptography-based scheme. Computer Networks, 46(2), 253-272.





access to further sensible information from the dataset. To anonymize domains, we will use the z-anonymity techniques, proposed in our previous work<sup>5</sup>. The idea at the base of z-anonymity is to release an attribute (a domain in our case) about a user only if at least z - 1 other users have presented the same attribute in a past time window  $\Delta t$ . Otherwise, it is blurred. In the following image, we report a graphical example of the z-anon concept with z = 3: a tuple is released only if other z - 1 = 2 different users have exposed the same attribute-value pair in the previous  $\Delta t$ . A complete description of the algorithm is available in the paper referenced above.



*Figure 26: Z-anonymity example* 

**Profile Creation and Analysis:** We use the Data Knowledge Extraction (DKE) PDK module to create users' profiles from the anonymized dataset. In a nutshell, the DKE learns relations among the different hosts visited by the users, allowing the profiling of users without having huge amounts of data. For reference on the complete behavior of the DKE we refer the reader to Gonzalez et al.<sup>6</sup>. We run the DKE module multiple times on the same dataset anonymized at different levels. Indeed, we test different levels of *z* (see the above paragraph for an explanation of the *z* parameter). Intuitively, *z* regulates the tradeoff between data utility and privacy. A large *z* filters out a large portion of the queries, providing strong privacy but leaving few information in the data. A small *z*, allows also infrequent domains to be released, reducing the privacy, but leaving a richer dataset. Here, our goal is to understand to what extent we can create useful user profiles using anonymized data.

#### Analysis of the results:

<sup>&</sup>lt;sup>5</sup> Jha, N., Favale, T., Vassio, L., Trevisan, M., & Mellia, M. (2020, December). Z-anonymity: Zerodelay anonymization for data streams. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3996-4005). IEEE.

<sup>&</sup>lt;sup>6</sup> Roberto Gonzalez, Claudio Soriente, Juan Miguel Carrascosa, Alberto Garcia-Duran, Costas lordanou, and Mathias Niepert. 2021. User profiling by network observers. In Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '21). Association for Computing Machinery, New York, NY, USA, 212–222. https://doi.org/10.1145/3485983.3494859







Figure 27: Cosine Similarity among profiles for different values of Z

The above figure quantifies how profiles change when data is anonymized. Each box reports the distribution of the similarity of a profile anonymized with a given z, with the original profile - see that when z=1, we are considering original profiles, thus similarity is 1. We measure similarity using the cosine similarity over the profile vector. The figures show how profiles get different with increasing anonymization. However, even with z=100, median similarity is around 0.97.



Figure 28: Effect of the Z parameter per interest

In Figure 28, we show how anonymization impacts the discovery of users interested in particular topics. The figure compares how similar is the set users interested in a topic obtained with a certain z with the set obtained from the original trace. With z=1, the similarity of the set, measured with the Jaccard Index between the sets, is 1. Increasing z leads to more imprecise sets of users. For some categories (e.g., Home & Garden) the accuracy remains high, while in other cases (e.g., People & Society) the sets are very different when using an anonymized trace.





In conclusion, with this demonstrator we have assess the viability of both the privacy preserving analytics and the User profiling modules in the network setting. In the one hand, the collection of data was possible using "in house" technology, without the need of installing new costly equipment. Moreover, the User profiling Algorithms demonstrated to be fast enough for the analysis of the data. On the other hand, we demonstrated that, even when applying a big anonymization factor (i.e., Z=50), the profiles constructed are similar. Thus, the application of Z-anonymity does not severely harm the usability of the data.

# 4.3.2 Location History demonstrator: Location data valuation

PIMCITY's Data Valuation Tool from the Users' perspective provides tools for data buyers to know what the value of data is for their specific tasks, without giving them access to such data. The valuation framework, which is the core of the tool, provides a series of abstractions to plug the buyers' models in, and automate such processing. This internal module of the PDK will be called by the TE to:

1) Select data, i.e., provide buyers with a hint of how valuable the information of a user is for a particular campaign, what we call try-before-you-buy (TBYB),

2) Calculate a fair breakdown of data transaction charges among users, rewarding each one proportionally to the criteria used to target them by specific buyers

Functionality 1:

their specific task, leading to a close-tooptimal purchasing process performance

Functionality 2:

Let buyers test data from individuals for Once a transaction is closed, let the TE reward users according to the value they bring to the specific task, i.e., according to the criteria they used to select them.



See: "Try Before You Buy: A practical data purchasing algorithm for real-world data marketplaces"

See: "Computing the Relative Value of Spatio-Temporal Data in Wholesale and Retail Data Marketplaces"

Figure 29: Functionality of the DVTUP shown in the demonstrator

In this demonstrator, we show the feasibility to use the PDK with real location data that can be collected by telcos and made available through a PIMS. In particular, we show how this module of PIMCity would help marketing companies target users that have consented to the PIMS using their location data for that purpose. As the figure below shows, we will use





preloaded models that work on pre-processed fully-anonymized *real* data from one of the telcos participating in the PIMCity Consortium to show the functionality of the module, and we will use Swagger to invoke the DVTUP API.



Figure 30: Architecture of the telco location history demonstrator

More specifically, we have pre-processed and cleaned data of the base stations used by more than 0.5M users of a telco, aggregated at base station and with 4-hour timeframes. For selecting the right users to target marketing campaigns, users have been ranked based on the number of times they were spotted during lunch or dinner time in the surroundings of three popular malls in the city of Liverpool. We have assumed that the more time users spend close to the mall during lunch and dinner time, the more valuable they are for restaurants hosted in that mall.

At the time of selecting the right users, the tool is called by the Trading Engine (TE) to select users based on the proximity to a given location. For example, the PIMS could provide the potential buyer with a table containing the number of times an anonymized individual was spotted close to the mall in the last week. That way, the privacy of users will be preserved while the TE helps buyers make a decision about how much and whose data to purchase. This is what we call try-before-you buy. The figures below plot histograms of such tables for the three malls, each of which is a point of interest (Poi or Pol) in this study.



Figure 31: Histograms summarizing data selection results

The histograms allow to differentiate casual from frequent users. Similarly, combining information of proximity to different malls helps target users of more than one store of a single brand, or "mall lovers", in this case 6K people that usually are present during lunch or dinner time in these kinds of places.



12K frequent Pol#3

Figure 32: Users found to be present in different malls

Once a transaction is closed, the marketplace will need to divide the resulting payment among the users that contributed data for that specific transaction. Such payment division is often equitable nowadays, but the value that buyers attain to users is not necessarily the same. The DVTUP provides functionality to carry out this payment division according to the criteria that buyers used to select the data. In this case the number of times a user was present close to the point of interest during lunch or dinner time in a given period. For that purpose, the trading engine will trigger what we call a valuation task. Should the Trading Engine decide to use this functionality, users that spend 4 days a week having lunch in one of these malls and whose information was sold to the marketing company would receive twice the reward for the transaction than those that only spend 2 days a week.

The source code is available at Gitlab (<u>https://gitlab.com/PIMCity/wp3/dvtup</u>, see readme.md for more information). It includes an update of the value-based data valuation framework, and the following new files:

- Sample data to test the application (Sample Telco Location Synthetic Data.zip)
- A Python script used to pre-process sample data and create the database resembling the PDS (TelcoData.py)
- New telco data valuation models (TelcoDataValuation.py) implementing the valuebased data valuation framework and responsible for carrying out the location data valuation.
- And a Jupyter notebook (TelcoDataDemonstrator.ipynb) showing the functionality of the tool in a local environment.

In conclusion, the demonstrator shows the feasibility of using the functionality in the PDK to support "human-centric" location-based targeting use cases. Buyers would be able to buy data or cast ads to users close to points of interest of their choice, and who have consented to the PIMS using their data for such commercial purposes. Moreover, using the same anonymized user information, the DVTUP allows the Trading Engine to reward users in proportion to their utility they have for the customer, in this case proportionally to the number of days they were spotted close to the corresponding Poi.





# 5 Conclusions

This document presents the final design and evaluation of the different pilots performed during the PIMCity project. The goal of the pilots is to demonstrate the ability of the different technologies developed during the project in real scenarios. To this end, three different pilots has been designed, focusing on three different aspects.

The first pilot is devoted to the continuous testing of the PDK module. This document updates D1.4 in this aspect. In this pilot, the main focus has been on testing the capacity of the different components individually. The different SMEs participating in the PIMCity consortium has tested and benchmarked different components in their daily activity. Moreover, this pilot has been performed in an Agile fashion, providing feedback to the developers in order to adapt the components to the real needs of the companies.

The second pilot has been designed to test the PIMCity solutions in a B2B scenario. In particular, we explored the possibility of integrating the concept of Transparency Tags in an already existing product for web protection, *Ermes for Enterprise*. The demonstrator has involved 6 companies in different industrial categories and with different size. All companies welcome the idea of TTs and feedback suggests they could be made accessible to final users of the products (e.g., companies employees) as a novel means to create awareness on privacy and security.

Finally, the main pilot of the project has demonstrated the usage of the technology in a B2C scenario. This pilot focused on the assessment of EasyPIMS, the PIMS created by integrating the components of the PDK. This pilot included different demonstrators including data from thousands of real users and real data buyers.

The first demonstrator focused on the interaction with final users. More than 3400 participated by registering in the platform and sharing their data. The second demonstrator enrolled companies that wanted to buy data from the platform. Finally, data from more than half a million users was used to demonstrate the ability of the components to handle the huge amounts of data present at the telcos.

The three pilots together present a perfect assessment of technology developed during the project, from the individual components, ready to be used by different companies, to integrated solutions of interest both to companies and final users.