Persistent Personal Data Vaults Empowering a Secure and Privacy Preserving Data Storage, Analysis, Sharing and Monetisation Platform

D6.5 Final Evaluation and Impact Assessment Report

Editor(s)	Shaun Topham
Lead Beneficiary	Assentian
Status	Final
Version	V1.00
Due Date	30/04/2023
Delivery Date	08/05/2023
Dissemination Level	PU

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871755.

Project	DataVaults – 871755
Work Package	WP6 Multi-Layer Demonstrators Setup, Operation and Business Value Exploration
Deliverable	D6.4 Pilots Evaluation of the Beta Platform Version
Editor	ASSENTIAN Shaun Topham
Contributors	ASSENTIAN- Ilesh Datani
	Suite5 –Giorgos Bikas
	Suite5 - Vasilis Galanos
	Suite5 - Sotiris Koussouris
	Prato – Paolo Boscolo
	Prato – Elena Palmisano
	MIWenergia - Borja Molina Rios
	MIWenergia – Ana Garcia Garre
	Andaman7 - Vincent Keunen
	Andaman7 - Sebastien Hannay
	Olympiacos - Maria Dolianiti
	Olympiacos - Christina Tsiligkiri
	Piraeus -Michail Bourmpos
	Fraunhofer - Yury Glikman
	ETA - Marina Cugurra
	ATOS - Iván Martínez
	ATOS - Tomás Pariente
	Maggioli - Nikos Achilleopoulos
	Maggioli - Sotiris Athanassopoulos
	Unisystems- John Kaldis
Reviewers	Suite5 – Sotiris Koussouris
	Andaman7 - Andrea Venegas

Abstract	This deliverable reports on the Evaluation of the Final Version of the DataVaults Platform, in accordance with the evaluation framework which was set out in D6.1.
Disclaimer	The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. © Copyright in this document remains vested with the DataVaults Partners

Executive Summary

This deliverable marks the final use of the Evaluation Framework as set out in D6.1 having been previously used to evaluate the alpha and beta versions of the DataVaults Platform.

The deliverable reports on the final evaluation for the project and how it has responded to stakeholder input.

It also checks that all the identified preparation work had been successfully carried out at the demonstration sites, to smoothly proceed through the final phase of the project, laying the foundations for exploitation, discussing in detail all the work performed in the different demonstrators of the project, throughout its duration.

The Evaluation Framework also emphasised “evaluation as a management tool” and in this capacity, we revisited our “Theory of Change” for DataVaults and took into account insights into how the project should have evolved in the final six months of the project, in this final evaluation.

The technical acceptance testing showed that the overall performance of the system and the experience enjoyed by the different user groups was of high standards, and improved during the project, as the platformed matured and new features and improvements were added to it.

The Legal, Ethical, Security, Privacy and Trust Evaluation demonstrate that both the ethics-and-privacy-by-design-and-by-default approach, as described in D9.2, and the requirements elicited in D2.1 and enriched in D2.3, have been properly followed by the technical team of during the design and development of DataVaults and by the demonstrators during the validation phase. In this regard, the DataVaults Ethics Advisory Board’s opinion confirmed that the consortium has worked so that the first version of the platform and its demonstrators reach the necessary conditions for legal, ethical and user friendliness, though further progress can still be made to reach an even better balance between legal / ethical constraints and user friendliness / user comprehension. This will be a gradual process that can be expected from future production systems. The same finding was also common to the TAM Questionnaire with the Data Owners. The outcomes of the EDPIAs and the fulfilment of the requirements regarding the DataVaults technology elicited in D2.2 and D2.3, confirm that the DataVaults platform, its app and the other technological assets generated in this research are legally compliant and ethically sound and give rise to a trusted, secure privacy-friendly and citizen-respectful data sharing environment. Feedback has been gathered from a broad community, based on the audience targets identified in D6.1.

The overall conclusion which was reached was that despite the impact of Covid and the required additional months to complete our work, there is a clear way forward for a concrete exploitation of DataVaults, guided by the lessons we have learned through the demonstration process.

Table of Contents

List of Figures

List of Tables

Terms and Abbreviations

API	Application Programming Interface
APP	Application
B2B	Business to Business
BDVA	Big Data Value Association
CO2	Carbon dioxide
CRM	Customer Relationship Management
D	Deliverable
DAA	Direct Anonymous Attestation
DoA	Description of Action
DLT	Distributed Ledger Technology
DPO	Data Protection Officer
DTU	Danish Technical University
EDPIA	Ethics and Data Protection Impact Assessment
EDPS	European Data Protection Supervisor
EFFRA	European Factories of the Future Research Association
EC	European Commission
EL	Ethical and Legal
EU	European Union
FAQ	Frequently asked question
GDPR	General Data Protection Regulations
GPS	Global positioning system
IEEE	Institute of Electrical and Electronic Engineers
IFIP	International Federation for Information Processing
IoT	Internet of Things
ISACA	Information Systems Audit and Control Association
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MVP	Most Valuable Product
ODI	Open Data Institute
PIA	Privacy Impact Assessment
REST API	Representational State Transfer
SME	Small and Medium Enterprise
SoTA	State of the Art
TOC	Theory of Change
TPM	Trusted Platform Module
UI	User Interface
UX	User Experience
WP	Work Package

1 INTRODUCTION

1.1 INTRODUCTION

"*DataVaults aims to deliver a novel framework and architecture that leverages personal data, coming from diverse sources, to help individuals construct their unified personal data hub, collect at a single point all of their personal data in a secure and trusted manner, and retain ownership and control on what to share and with whom, also receiving compensation for the artefacts they place at the disposal of other third parties."* [1]

This document, as set out in the DoA, covers: "Documentation of the demonstrators' operation and execution consolidating the input of Tasks 6.3-6.7." This report will evaluate the final version of the platform. It is the final activity from T6.8-Demonstrators Evaluation and Impact Assessment, which commenced at M19 and ran through to M40.

The exhaustive evaluation framework established in D6.1 has been utilised through the alpha and beta phases of the technical development, and throughout the period in which the demonstrators have been operational. The demonstrators being:

- Sports and Activity Personal Data (OLYMPIACOS)
- Strengthening Entrepreneurship and Mobility (PIRAEUS)
- Secure Healthcare Data Retention and Sharing (ANDAMAN7)
- Smart Home Personal Energy Data (MIWENERGIA)
- Personal Data for Municipal Services and the Tourism Industry (PRATO)

Extensive data collection, regarding the experience of the demonstrator partners with the DataVaults platform, was conducted throughout the running phases. The data collection methods met the guidelines of the predefined evaluation framework and was in accordance with the evolving Evaluation Plan (Task 6.2), to ensure the high quality of the feedback gained and consistency of the evaluation activities.

The Evaluation Plan was continuously updated as the project progressed as "*An effective evaluation plan is a dynamic tool, or a 'living document', that should be updated on an ongoing basis to reflect changes and priorities over time.*"[2] The evaluation framework was a useful tool, which lead to valuable observations and conclusions about the viability and the sustainability of the DataVaults platform.

D6.1 set out to describe how we were to evaluate both the pilots and the project as a whole. It provided the "Documentation of the evaluation framework and validation methodology, defining the various practices for recording feedback from the demonstration activities and including a set of test-cases to be executed by the demonstrator partners."

[1] DataVaults, "DataVaults DoA. Section 1 3.1, Project Concept."
[2] United Nations Environment Agency, "evaluation-criteria-and-ratings," https://www.unenvironment.org/about-un-environment/evaluation-office/our-evaluation-proach/evaluation-criteria-and-ratings., [Accessed:21 Jan 2021]

The demonstrations themselves commenced at M19 (July 2021). This document builds upon the initial evaluation of the alpha version of the platform, presented in D6.3 at M24 and D6.4 which evaluated the beta version at M30.

The earlier evaluations enabled us to:

- Check that the technology being developed was going in the right direction and should work according to the requirements established in WP1.
- Continue validating that the assumptions made about the project were correct.
- Check all the conditions were right and progress has been performed in alignment (or with small deviations) to the original execution plan, including whether the needs and requirements of the stakeholders and of the personal data market are being taken into account.

This final evaluation will return to the basic overall evaluation strategy set out in D6.1 and respond to all the aspects identified at that stage as being relevant to the success of the project, embracing feedback from stakeholders.

The beta version was made available in M24 of the project (end of December 2021). It included all the "basic" functions, which were specified in D5.2, following the availability of the mock APIs and allowed stakeholders to perform real-life tests with the platform becoming close to a stable prototype that can perform most of the core operations relevant to data collection and sharing.

Following on from this, with the availability of the final platform, the consortium as in a position to fully evaluate the overall operation of DataVaults, based on the v0.50 and v1.0 releases of the platform, which included improvements over the beta version (both at the UI as well as at the backend level), and also some new additions to existing features to make the platform more complete and easy to use.

As in D6.3 and in D6.4, the legal, ethical, security, privacy and trust aspects of the DataVaults technology (Cloud Platform and Personal App) and of the demonstrators where such technologies were validated were evaluated against the requirements originally set out in D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management" and then refined and updated in D2.3 "Updated DataVaults Security Methods and Market Design", as well and against WP10 "Ethics Requirements". The assessment focused, also in this final evaluation, on:

- Layer I: the DataVaults demonstrators' processes and operations (Sect. 2.2), including the ethical procedures and the preparatory activities in view of the interaction with the citizens.
- Layer II: DataVaults technology (platform and app) (Sect. 2.3.), encompassing the assessment of the components, tools, and services under development in the project and which were validated (as in the previous evaluation iteration) in the demonstrators. This layer mainly relied on the Ethics and Data Protection Impact Assessments (EDPIAs), though in this final evaluation iteration this layer also

encompasses some key findings from the data owners in the demonstrators' environment, captured thorough dedicated questions in the TAM Questionnaire.

Based on these two layers, the final evaluation of the Legal, Ethical, Security, Privacy and Trust aspects was performed, and its findings are reported in this document under Section 2. Furthermore, the lessons learnt and key takeaways towards legal compliance and ethical soundness of DataVaults Technology are reported in the Section 5 of D6.6 "DataVaults Scaleup Roadmap and Key Takeaways".

1.2 DOCUMENT STRUCTURE

The document reflects the structure of the previous deliverables, which evaluated the alpha and beta versions. Essentially it can be seen as having three sections.

The **first section** poses the question as to whether all the envisaged preparation work has been successfully carried out. This will embrace a wide range of questions, including:

- Whether the expected progress, as set out in the DoW for the project has been achieved, at all the demonstration sites in accordance with what was planned in D6.2?
- Have the conditions for having an ethical and privacy situation in keeping with the prescriptions of WP2 and WP10 been met?
- Are the identified interactions with citizens capable of being carried out at each of the demonstration sites?
- Was there sufficient interaction with all the identified stakeholders?

This section is covered by chapter 2 to chapter 4.

The **second section** will concentrate on the evaluation of the technical work from WP3 and WP4, with WP5 acting as the conduit linking the work. The technical approach will be covered and the results from the technical evaluation of the final version will be brought together. For this purpose, WP5 created specific tests to ensure quality of the code and the required test flows to cover the main aspects of the integration issues.

This section is embraced by Chapter 5.

And the **third section** will embrace progress made at a holistic project level and will check whether the assumptions we have made in D6.1 and in the overall project remained valid at the end of the project, drawing attention to where any improvements were made. As referred to above, the whole evaluation framework and evaluation process was a dynamic tool and a living document to be studied and to be updated on an ongoing basis, to reflect changes and priorities over time. In this instance, it will also be augmented by the deliberations around the "lessons learned process" and in establishing the pathway for the third parties who will be deploying the platform, set out in D6.6 DataVaults Scaleup Roadmap and Key Takeaways [M40] This deliverable provides the Documentation and lessons learnt from the DataVaults project, constituting methodological adoption guidelines for the utilisation of the platform.

Questions posed include:

- Have we identified any changes or new requirements from any of the key stakeholders?
- Are we learning lessons regarding our “offering” or is there any new emphasis required in terms of our roadmap?
- Did the DataVaults “Theory of Change” remain true?

This section is embraced by Chapter 6, whilst with Chapter 7, we will draw out the conclusions which we can make from the evaluation of the final version of the platform.

Finally, Appendix 1 returns to the table which was produced for D6.1, which laboriously captured all imaginable questions which could be asked of the DataVaults project, to ascertain its overall progress. This resulted in a “broad brush-stroke” overview of the evaluation of the project complementing the detailed work carried out within the final evaluation of the project. Whilst Appendix 2 includes the DataVaults Theory of Change diagrams for reference.

2 LEGAL, ETHICAL, SECURITY, PRIVACY AND TRUST: FINAL EVALUATION

2.1 OBJECT OF THE LEGAL, ETHICAL, SECURITY, PRIVACY AND TRUST EVALUATION

As in D6.3 and D6.4, this section dwells upon the evaluation of the legal, ethical, security, privacy and trust aspects of the DataVaults technology (platform and App) and of the demonstrators where such technologies are going to be validated and assessed against the requirements elicited in the previous stage of the project. In particular, regarding the legal, ethical, security, privacy and trust requirements, they have been originally set out in D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management" and then refined and updated in D2.3 "Updated DataVaults Security Methods and Market Design". Such documents also outlined the surrounding challenges regarding such topics, as well as providing guidelines and hints aimed at supporting their operationalization, both during the execution of the project and, to some extent, in the post-project adoption of DataVaults outcomes. In particular, the assessment focused, in this final evaluation, on:

- **Layer I: the DataVaults demonstrators' processes and operations** (Sect. 2.2), including the ethical procedures and the preparatory activities in view of the interaction with the citizens. The demonstrators themselves need to adhere to the Ethical and Legal requirements specifically set for them either within WP2 or in WP10 "Ethics Requirements", such as the need to follow ethical procedures (consent procedures, recruiting procedures, etc.) and to use adequate tools (such as the refined consent form and information sheet, the use of suitable inclusion/exclusion criteria for the involvement of volunteers in the demonstration activities, etc.).
- **Layer II: DataVaults technology (platform and app)** (Sect. 2.3.), encompassing the assessment of the components, tools, and services under development in the project and which are validated (as in the previous evaluation iteration) in the demonstrators. This layer relied on the Ethics and Data Protection Impact Assessments (EDPIAs). One EDPIA for each demonstrator has been conducted, as a key tool for the assessment of the legal, ethical, security, privacy, and trust aspects of DataVaults technology in the different contexts. In addition, the Layer II was enriched with findings from the data owners in the demonstrators' environment. This assessment focused on the citizens' perspective on aspects impacting the legal and ethical dimensions of the DataVaults technology (platform and app). The citizens' feedback was gathered with dedicated questions in the TAM Questionnaire, together with the other non-functional requirements questions.

Based on these two layers, the final evaluation of the Legal, Ethical, Security, Privacy and Trust aspects was performed, and its findings are reported in this Section. The Layer II in the past iterations was particularly useful for the future development work and for the operationalization of the ethics-and-privacy-by-design-and-by-default approach. On the other hand, the evaluation addressed by Layer I can be useful, besides for EC to monitor our compliance with H2020 ethical standards, also for the post-project phase, for instance through the elaboration of lessons learnt or blueprints for DataVaults' uptake and operation in real-life environments. In this regard, the lessons learnt and key takeaways towards legal

compliance and ethical soundness of DataVaults Technology are reported in the Section 5 of D6.6 “DataVaults Scaleup Roadmap and Key Takeaways”.

2.2 ETHICAL PROCEDURES AND INTERACTION WITH CITIZENS (DEMONSTRATORS)

Evaluation Questionnaire on Ethical and Legal Aspects	
#1 Your reason for using personal data	
What is your primary purpose for collecting and using personal data in DataVaults? For which use cases/scenarios do you need them? Are you replacing another product or service as a result of DataVaults?	
Olympiacos	Our primary purpose for collecting personal data and demographics data is to improve the collection of data relevant to the fan base, such as location data, fan-based activity data, and preferences from clubs' fans and members. In addition, we collected biometrics data from athletes to test the option to have a central way of collecting such data towards better management of statistical reports and better management of their results from ergometric tests and medical examinations. These are reflected in our two scenarios. With DataVaults platform we complement existing applications that hold very generic fan base data, as well as we aim to replace the manual way of holding athlete’s data in paper/digital documents.
Piraeus	Our primary purpose for collecting personal data is to: a) improve mobility around sport venues at times of events. b) enhance the local commercial market, using citizens commercial profiles. c) provide tailored cultural and touristic experiences to tourists and citizens. These activities are depicted in the three scenarios included in the demonstration activities of DataVaults.
Andaman7	The first main reason to collect data is to use them to provide new features to our users and improve the attractiveness towards our app to bring new users. The second reason is to be able to help build a better health system by sharing part of such data to B2B partners (with user consent) such as medical research companies. Data collection is needed in both scenarios. Whilst the first

	scenario mainly concerns the collection for B2B partners, the second scenario aims to improve the Andaman7 application. DataVaults will improve our existing product.
MiWenergia	Our main purposes for collecting energy consumption data are to be able to recommend energy-saving tips, design PV installations and generate energy demand prediction models. These are reflected in our three scenarios. Data collection through DataVaults can help to design more targeted marketing campaigns and services. In addition, we now use an external application to predict our client's energy demand, which could be substituted with our own model trained with data from DataVaults.
Prato	As Prato Municipality, the primary purpose is to access citizens' personal data for improving mobility services and contributing to enrich the cultural offer in the city, together with the involved museums. Use cases are related with mobility, cultural offer and exchange of personal certificates. In the case of mobility and culture, DataVaults might substitute current approaches in the acquisition of citizens' personal data, based on the purchase from big data players.
#2 Recruitment procedures	
Have you already followed the recruitment procedures, including the inclusion/exclusion criteria, described in D10.1 to identify/recruit research participants?	
Olympiacos	Questionnaires played an important role in recruitment procedures. During the past period (June – September) our fans and members renew their annual subscription, and this help us to receive their answers. In addition, official invitation by email (newsletter) was send to selected group our fans and members to participate on the DataVaults app, following the criteria for onboarding of D10.1
Piraeus	Questionnaires are ready to be deployed to all the people involved in the pilot. All participants are volunteers. A small group of people within the Municipality have been identified in order to test the questionnaire module of the DataVaults app. These people are mostly employees of the Municipality of Piraeus and friends/family, all citizens of Piraeus

	and connected based on their profiles to the relevant pilot. All of them are volunteers.
Andaman7	We followed the recruitment procedure as described in D10.1. The first phase (contacting only users who answered the Survey on citizens' perspective, expectations, needs and concerns on Personal Data Sharing) gave no result so we quickly shifted to the second phase. However, given the time frame left to run the demonstration activities, inclusion prioritisation was not followed as closely as expected during the redaction of 10.1 to be able to start those activities in time to have results. We still applied a quick check on all candidatures before including them in the demonstration
MiWenergia	Yes, we have followed every criterion described in D10.1. Every participant has signed the correspondent informed consent and we have informed every participant that their participation is voluntary. The process is finished.
Prato	As far as the certificate scenario is concerned, contacts have been established with the fiscal support centre as DS to test the procedure by the end of March. As for the DO side, a small group of users have been established so far, able to test the procedures related with the certificate scenario and the answer to questionnaires related with mobility and culture. They will be involved to check the scenarios once the technical tools are fully operational, including the connection with the mobile app providing user's routes.

#3 Informed consent procedures	
Have you already followed the informed consent procedures described in D10.1 for the participation of humans? Have you already introduced any changes to simplify the informed consent and information sheets inserted in D10.1, in order to make them more intelligible to the research participants and/or to adapt them to the online environment?	
Olympiacos	Informed consent procedure is now fully integrated in our registration process of the DataVaults app.
Piraeus	The informed consent procedure is accessible through the DataVaults platform.

Andaman7	The informed consent procedures are followed as described in D10.1. During the Andaman7 onboarding, as we are on mobile devices, a simplified procedure is presented to the user. We present DataVaults with three short and simple texts then provide links to the full informed consent and information sheets provided in D10.1 without simplification. The user can then choose to copy links and consult documents on a computer rather than on the smartphone. Conditions should be read and accepted but we don't have any check on the consultation of the complete versions of documents
MiWenergia	Yes, we have followed all the informed consent procedures. We have not introduced any change, but the informed consent has been integrated in the DataVaults platform registration process.
Prato	The informed consent procedure is published and accessible through the DataVaults platform when the user selects the participation in the Prato pilot.

#4 Security and privacy-preserving measures	
Have you already taken the measures for data collection, handling, storage, protection, retention and destruction as described in D10.2, as well as for managing the rights of the users?	
Olympiacos	Yes, all data collected has been following the measures described in D10.2, as it is on the DataVaults platform.
Piraeus	We handle and store all the information as described in D10.2.
Andaman7	All security and privacy-preserving measures were already integrated in our Andaman7 platform/app before the DataVaults project.
	Yes, we have handled and stored correctly all the information. We have not yet destroyed any documents because no user has claimed his right to do so.

MiWenergia	
Prato	We carried out an impact analysis based on the platform details provided by technical partners. Safety and privacy-preserving measures are in place according to technical deployment, currently no user has claimed right to destroy content.

#5 Profiling	
In case in your demonstrator profiling is expected to occur, have you already informed or are you going to inform the data subjects on it, including also on its possible consequences, according to what planned in D10.2? And in such a case, have you already taken the mitigating measures described in D10.2 to safeguard their rights?	
Olympiacos	Information on profiling is already included in the informed consent form to be used.
Piraeus	Information on profiling is already included in the informed consent form to be used. No profiling has taken place so far.
Andaman7	Our scenarios don't involve any profiling.
MiWenergia	Yes, we have implemented all the measures required. Information on profiling is already included in the informed consent form.
Prato	Information on profiling is available in the informed consent form accessible to real users by entering the Prato pilot on the platform. No real profiling has taken place so far.

#6 Special categories of personal data
Have you planned and/or implemented proper safeguards for the special categories of personal data, if any in your demonstrator? Have you planned and/or implemented adequate measures to ensure that such special categories of personal data will not be used beyond the original purposes?

Olympiacos	All safeguards for the special categories of personal data were already integrated in our CRM platform before DataVaults project. Only anonymised/statistics data will be collected and stored in our premises.
Piraeus	We have not planned any safeguards or measures as we are not collecting/processing special categories of data.
Andaman7	All safeguards for the special categories of personal data were already integrated in our Andaman7 platform/app before DataVaults project.
MiWenergia	We have neither planned nor implemented any safeguards or measures as we do not have special categories of data.
Prato	We carried out an impact analysis based on the platform details provided by technical partners. Should there be any need to apply such measure, we will proceed accordingly.

#7 Participation to the DataVaults Ethics Board	
Has a representative of your demonstrator taken part to the activities of the DataVaults Ethics Board (meetings, opinions on deliverables relevant to ethics and privacy, planning of adequate countermeasures, etc.) and brought the specificities of your demonstrators to the attention to the other members, as described in D10.3?	
Olympiacos	Yes, we have attended the Ethics Board.
Piraeus	Yes, we have attended the Ethics Board.
Andaman7	Yes, specificities were communicated during ethics meetings and through deliverables.

MiWenergia	Yes, we have attended all the Ethics Board.
Prato	Yes, one representative is member of the Board and actively takes part in the board discussions, for example as far as compensation methods for public administration are concerned.

#8 Positive impact on people	
Have you considered the positive effects on people of the implementation of your demonstrator? *Are you making things better for society? How and for whom?* *Which individuals, groups, demographics or organisations will be positively affected by it? How?* *How are you measuring and communicating positive impact? How could you increase it?*	
Olympiacos	We have considered the positive effects for data owners and data seekers. The club, using DataVaults is expected to provide marketing campaigns and useful analytics. Moreover, we assume that the demonstrator will contribute to arising members' awareness on personal data management risks and will empower users in maintaining the control on their data sharing.
Piraeus	All scenarios are either targeting the improvement of the life of citizens and tourists (mobility and tailored touristic/cultural experiences) or the enhancement of local economy (empowering local shop owners and entrepreneurs). We are planning surveys in order to measure the expected positive impact, as described in the AF of DataVaults.
Andaman7	Our main objectives are improving the app to help users to take care of their health and ease collection of data for companies that build a better healthcare system for the future. We then aim to have a positive effect on our users (patients), healthcare companies and ultimately the whole healthcare system. The KPIs we are planning to measure should help having a sense of those positive impacts.
MiWenergia	Yes, we will try and provide users with useful energy efficiency tips while educating them in awareness of renewable sources and efficiency use of resources. This will not only benefit the individual

	but will also be good for all society because it can reduce the impact on the environment.
Prato	Positive effects might arise both for all groups of citizens and data seekers (municipality, cultural institutions, and fiscal service providers) in the improvements of public and cultural services and in the facilitation in the management and sharing of personal certificates. Moreover, we assume that the demonstrator will contribute to raising citizens' awareness on personal data management risks and will empower users in maintaining the control on their data sharing.

#9 Negative impact on people	
Have you considered the possible negative effects on people of the implementation of your demonstrator? *Who could be negatively affected by it?* *Could the way that data are collected, used or shared, cause harm or expose individuals to risk of being re-identified?* *How are limitations and risks communicated to people (both people whom the data are about, and people potentially impacted by their use)?* *What steps can you take to minimise harm?* *How are you going to measure, report and act on potential negative impacts of your demonstrator?*	
Olympiacos	We do not foresee any negative impact on people, related to the scenarios and the type of data collected.
Piraeus	We do not foresee any negative impact on people, related to the three scenarios and the type of data collected.
Andaman7	The transparency policy described in previous deliverable is still valid and aims to reduce stress and fears related to sensitive data sharing. We do not foresee any other negative impact. A transparent communication and quick responses to any received feedback via a dedicated employee has been put in place from the day 1 of demonstration activities.

MiWenergia	We do not consider the users will face negative effects from our collection and use of their data. They will be always aware of what kind of data are sharing with us and in which manner (anonymised, encrypted, etc) and we will be supporting them to ensure transparency. If any user reports any harm, we will take actions to fix and reduce any negative consequence.
Prato	We are not foreseeing any negative effects on people and the technical solutions put in place in the DataVaults tools will reasonably prevent the possibility of individuals' re-identifications. Nevertheless, the survey tools that will be adopted in the open demonstration phase might possibly highlight negative impacts perceived by the users. Should this be the case, specific countermeasures will be put in place accordingly.

#10 People's feedback	
Are you routinely building in thoughts, ideas and considerations of people affected in your demonstrator? How? *What information or training might be needed to help people understand data issues?* *Are you going to establish proper interaction mechanisms with data subjects for communicating them their rights and notices, answering their questions and providing further information to them about the data processing?* *Are you going to use some specific functionalities of the DataVaults platform and app?*	
Olympiacos	People's contribution will be requested by pushing a questionnaire related to the demonstrator activities, directly through the DataVaults platform. In-house demonstration activities performed allowed to explain to individuals the concept and the way of operation of the platform.
Piraeus	We plan to ask for user's feedback via questionnaires, either standalone or through the DataVaults platform.
Andaman7	Spontaneous feedback through email and regular feedback inquiries through questionnaires/direct exchanges are collected. We also check regularly if and how citizens are able to complete steps of demonstration activities through data we can access. If any difficulty is spotted during the process, we contact all concerned citizens and try to help them and ease the process.

MiWenergia	We will ask for user's feedback via questionnaires through the DataVaults PersonalApp and externally. We are also available for them to contact us via email or phone if they have any issues.
Prato	People's contribution might be taken into account by activating some feedback process related with the demonstrator activities, for example through specific surveys. The topic of personal data sharing is increasingly important, and some specific communication and training actions could be carried out through social media and institutional channels. At the moment no specific communication tools are included in the DataVaults app/platform, but should this happen, we will consider this opportunity.

#11 Privacy-friendliness	
Do you perceive DataVaults as a privacy-friendly technology?	
Olympiacos	Yes, the DataVaults tools provide various means to protect individuals' privacy.
Piraeus	Yes, we consider DataVaults to be a privacy-friendly technology.
Andaman7	Yes, the platform still seems to be built with privacy at the heart of the project.
MiWenergia	Yes, the platform is built with privacy at the heart of the project.
Prato	Yes, this is essentially the purpose of the DataVaults tools.

#12 Societal Impact

From your point of view, what segments of society are expected to benefit most from DataVaults deployment?	
Olympiacos	The benefits provided by DataVaults are expected to be felt by a wide number of segments in society, since the topic of personal data management and sharing is currently increasing importance at all levels. Especially this could have an impact on youth, as they might reconsider the way they are sharing data, and the purpose of such actions, by understanding the paradigm of DataVaults
Piraeus	From our point of view, benefits are expected for both the data users and data seekers as described in DataVaults. Data users, in our case citizens and tourists, may benefit directly from the compensation mechanism and indirectly through the improvement of their everyday life (mobility, touristic/cultural experiences), while data seekers, in our case the municipality or the local trade association will benefit in terms of having an extra tool to acquire much needed information for strategic planning.
Andaman7	From the health point of view, the main segments that could benefit are both patients (to be empowered with their data and contribute to better care and research) and the medical research/medical care companies that run clinical trials. The DataVaults platform should bring a powerful tool that will ease recruitment and data collection for such trials. Given the fact that those are usually built to bring better health for everyone, the whole health system should ultimately benefit from it.
MiWenergia	From our energy point of view, people who are responsible for their homes and who are not very educated in energy efficiency and renewable sources will benefit greatly. Also, people interested in reducing their energy consumption and their environmental impact can be benefited from the DataVaults deployment. At the same time, if participants reduce their energy consumption or increase their consumption from renewable sources, it will have a positive impact on the environment, reducing CO2 emissions.
Prato	Potentially, the benefits provided by DataVaults are expected for a wide number of segments in society, since the topic of personal data management and sharing is currently increasing importance at all levels. Of course, since DataVaults is a technological solution connecting digital data sources, it will address those using digital devices and accessing digital services, who represent certainly an ever-growing part of the population.

Table 1: Evaluation Questionnaire on Ethical and Legal Aspects

2.3 EVALUATION OF THE DATAVAULTS TECHNOLOGY IN THIS CONTEXT

At the core of DataVaults is the provision of safe, secure, private, fair, legal and ethical mechanisms for handling personal data. Therefore, this aspect of the Evaluation Plan was designed to ensure that we conform to everything that has been promised in these respects. Table 2 below sets out how we propose to do this, indicating the metrics, success criteria and data collection process.

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Frequency of data collection	Final evaluation
Citizen experience					
C.1 Improving Privacy Risk Exposure Awareness for Individuals when sharing Personal Data. **KPI**	Starting point: zero	100% with introduction of new methods. Delivery of the Platform and of the Dashboard.	The privacy risk metrics dashboard will provide this information as opposed to the current situation where no such data is available.	From date of availability of dashboard, as appropriate	100% as each sharing action includes a risk index letting individuals know their exposure
C.2 Improve-ment of Individuals Knowledge on Personal Data Safe-guarding.	Did this facilitate risk situational awareness?	Individuals engaged on the platform will understand better how to share and safeguard their data. Show an increase.	Verified through the AS-IS vs. TO-BE evaluation in the demonstrators through surveys, interviews etc.	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners	The various communications and explanations on the DataVaults technology to the users contributed to the general understanding that their data are important to them
(For example, health data is important for good care) but also have value for others (for example my health data can be useful for clinical research). The training and demo sessions have helped the participants to understand the relevance of their personal data and the way in which their data can					

help to the development of new technology and services that are good not only for them as individuals but also for the society. The DataVaults general communication approach was probably more "positive" than the general communication by the press and social networks on "big tech is abusing your data, therefore yourself too". However, especially the demonstrators recognised that this is a global problem which is slowly being improved over the years. The DataVaults project, as well as the projects of our demonstrators (for example Andaman7 contribute to improve the general literacy of citizens on the importance of their personal data					
C.3 To overcome reluctance to personal data sharing services via DataVaults	Did the DataVaults system contribute to build trust and overcome the data owners' reluctance in sharing their personal data?	Individuals will be educated on how their personal data can be used, while also enjoy remuneration for this, resulting in a trust building. Use of service	Verified through the AS-IS vs. TO-BE evaluation in the demonstrators through surveys, interviews etc.	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners	Reluctance to share personal data is normal given all the abuses by some companies. The situation can only be fixed by gradually rebuilding the trust. GDPR is an excellent step in the right direction of setting strict guidelines for all stakeholders to adhere to trust generating principles. DataVaults contributed to rebuilding trust, also thanks to its training and demo sessions.
Did these facilitate both privacy and trust preservation and has there been an Improvement in Trust? And TOC1 Have we given improved control and awareness of how a citizen's data is shared and managed?			Observation from C1, C2 and C3.		Yes, the project activities were effective in this direction, though further work must be done
C.4 Increase of the value of personal data attributed back to owners **KPI**	Starting point: zero	100% with introduction of new methods. Currently no value is attributed to data owners, and with DataVaults this will change.	Verified through the DataVaults methodology and the AS-IS vs. TO-BE evaluation in the demonstrato	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners	Yes, though due to the short time that the pilot has run it was not possible to assess this extensively. Individuals receive the points that they feel that are fair

Are individuals receiving a fair share of their data value?			rs through surveys, interviews etc.		from the data they are sharing. It is an open market, so they have the "power" to decide for which amount of "money" they share their data.
C.5 To develop privacy metrics that are easy to understand for data subjects	Development of user-friendly privacy metrics, whose testing occurred in the final phase of evaluation	100% with introduction of privacy metrics and information to the data owners through easy-to-understand privacy notices	Verified through citizen testing and AS-IS vs. TO-BE evaluation in the demonstrators through surveys, interviews etc.	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners Privacy Measures to determine the susceptibility of data or a dataset to revealing private information. (https://arxiv.org/abs/1512.00327)	There has not been enough time for the demonstration phase to know if the metrics developed are easy to understand for individuals. Due to the short time that the demonstrator was run, we have no feedback on this. Despite some useful hints arose from TAM questionnaires, a more complex use (more datasets shared) will be required to receive accurate feedback.
C.6 Has the personal app been successfully used by individuals for storing, collecting and sharing data and what was their experience here?	N/A	100% with the use of the personal app by individuals. Currently individuals do not feel fully-comfortable in sharing their personal data with personal data platforms. With DataVaults this will change, thanks to its functionalities for user control and the fair, transparent share of the value of data (trust building and acceptance of DataVaults)	Verified through citizen testing and evaluation in the demonstrators through surveys, interviews etc.	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners	It has been used to some extent but the time for running the evaluation was limited. During the demo/training sessions it emerged that the citizens feel that further efforts should be done to make the platform easier to use and that there are many privacy and anonymization terms that they don't understand. Probably more time

					and more training sessions, with more time to simplify and address their feedback of the platform will have helped on this issue.
C.7 Was the personal app consent mechanism clear and well-received?	N/A	The consent form and information sheet already provided for each of the demonstrators will be refined and updated in order to adapt them to the online environment, the layered approach for consent management described in D2.1 and the indications coming from the EDPS's Guidelines 05/2020 on consent under Regulation 2016/679Version 1.1 (2020)	Verified through citizen testing and evaluation in the demonstrators through surveys, interviews etc.	Surveys and interviews as appropriate, in particular the TAM Questionnaire to Data Owners	Yes
Legal, Ethical, Security and Privacy aspects					
R.1. "Personal data platforms shall ensure respect of prevailing legislation and allow data subjects and data owners to remain in control of their data and its subsequent use."		Is there compliance with list of requirements identified in WP2?		Section 2.3.3 hereunder shows the state of achievement of the requirements set in D2.2 and D2.3, including compliance with the applicable legislation.	The design of DataVaults technology is taking into account the requirements set in WP2.
R.2. "… Conditions of use and practical arrangements of data sharing should be regulated."		The approach and key choices related to the Smart for the fair and secure personal data management have been defined in T2.3 and is under implementation within WP3, WP4 and WP5		Section 2.3.3 hereunder shows the state of achievement of the requirements set in D2.2 and D2.3, including compliance with the applicable legislation.	YES. The Data Sharing configuration service has been developed and is integrated for this purpose

R.3. Has there been a contribution to the Basis for Privacy, Ethics and IPR?	This will be part of Lessons Learned.	This has been reported in D6.6	N/A
Do we contribute to and are we aligned with:			
R.4. Ethics driven guidelines and privacy & security standards	Several data ethics guidelines have been central to the design and ongoing development of DataVaults technology, such as the EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act) Version 1.1 (March 2021), the EDPS Opinion 7/2015[3]), as well as an Open Data White Paper and ENISA's Privacy standards for information security.[4]	Final Assessment in D6.5	YES. The overall design and development of the DataVaults technological artefacts were aligned with the ethics-driven guidelines and privacy & security standards, as resulting in the WP3, WP4 and WP5 deliverables.
R.5. Fundamental Rights and Well-being	The design and ongoing implementation of the DataVaults technological artefacts are driven by Fairness & Privacy-by-Design-and-by-Default approach, enriched with the Protection Goals method, as described in D9.2 and in D2.1	Final Assessment in D6.5	YES. DataVaults technologies were directed to uphold European values, for respecting fundamental right of the data owners and promote their well-being and empower-ment, as resulting in the WP3, WP4 and WP5 deliverables
R.6. Is the personal app compliant with EU regulations and national laws?	D2.1 and D2.3 provide the identification and description of the regulatory sources relevant to the DataVaults components, personal App and cloud-based platform, together with a series of mandatory and recommended legal requirements to be met	Final Assessment in D6.5	YES. The design and development of the DataVaults technologies took into account the relevant legal sources and are compliant with them

[3] "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability" (2015

[4] https://www.enisa.europa.eu/news/enisa-news/privacy-standards-for-information-security

R.7. Did we provide an Ethics monitoring framework?	D9.2, Sect. 2 describes the overall ethical policy of the project. Such a Policy, besides describing the Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals Approach, comprises a snapshot of the ethical procedures and responsibilities, the definition of the oversight responsibilities (DataVaults Ethics & Data Protection Officer and DataVaults Ethical Board) and preliminary considerations for the elaboration of the comprehensive Ethical and Data Protection Impact Assessment methodology to be used within WP6, during the demonstrators' operations. Furthermore, D10.2 section 3.1.1 provides a comprehensive ethics risk evaluation table in relation to the pilot's activities. In addition, the Ethics Board has been set, including independent expertise, and is operating with strong interaction with the technical team.	N/A	YES
R.8. Did we successfully adopt novel trusted and security-by-design data mining, management, analysis and sharing techniques?	The approach and processes were outlined in T2.2, where also the key technological choices were taken.	In line with the approach and techniques set in WP2, these aspects were initially covered by D3.1 and have been further addressed in D3.2 and D3.3, as well as in the WP5 deliverables	YES. The final architecture and component facilitate this need
R.9. Have privacy analytic methods been provided and tested?	Details on the approach and procedures for data collection, storage, protection, retention, and destruction to be implemented by the DataVaults tools (app and platform) are and will be further provided in the technical deliverables released by the DataVaults consortium in WP2, WP3, WP4 and WP5.	Next WP3, WP4, WP5 deliverables	

Table 2: Overall evaluation of citizen experience and of Legal, Ethical, Privacy and Security aspects

2.3.1 Ethics and Data Protection Impact Assessments - Final release

This section reports the final release of the Ethics and Data Protection Impact Assessments (EDPIAs), elaborated for each of the DataVaults Demonstrators, with a strong collaboration among the demonstrator partners and the technical supporting partners, besides the overall technical team of the project. These EDPIAs also referred to the legal and ethical requirements set in D2.1 and D2.3 and considered the specific DataVaults technologies (like services, components) relevant to their context, the data lifecycle and each use cases scenarios, as well as their own privacy and security policies/practices.

In DataVaults, the EDPIA served as a tool for "the assessment of risks for individuals' rights, freedoms, and wellbeing, for ensuring compliance with the data protection law (GDPR and national regimes), and ethical mandates"[5]. As clarified in D6.3 and in D9.2, the assessment methodology regarding the risks for the personal data was conducted following the indications of Article 35 sec. 1 GDPR, taking into account the nature, scope, context and purposes of the processing operations in each demonstrator in view of evaluating their impact on the protection of personal data, to identify and reduce the data protection risk[6] and the likelihood of privacy harms to individuals, as well as to identify and put in place the appropriate technical and organisational measures to tackle with/mitigate such risks. For these purposes, the Consortium adopted a model inspired by the ISACA Model[7], mapping the fourteen ISACA privacy principles to the specific GDPR requirements.

The ethical dimensions of the personal data collection and processing and the assessment to what extent the principle of fairness has been operationalized in each of the demonstrator, this ISACA-inspired model was enriched with the Data Ethics Canvas, elaborated by the Irish ADAPT Centre for Digital Content Technology on the basis on the original Business Model Canvas by Alex Osterwalder. It consists in a useful tool giving a higher-level framework to develop ethical guidance that suits any context and to assess the ethical implications of any project, thereby allowing to be more trustworthy with data processing.

More information on the DataVaults EDPIA's model can be retrieved in D6.3, sect. 2.3.1.

The EDPIA was conducted for each demonstrator through a questionnaire, comprising elements coming both from the ISACA Model and from the Data Ethics Canvas.

The following paragraphs contains the consolidated version of the DataVaults EDPIAs, notably one for each demonstrator.

2.3.1.1 *Demonstrator #1 – Sports and Activity Personal Data*

Introductory questions		
		Responses
1	Nature, scope, context and purposes of the processing (Recital 90)	We collected activity data and personal preferences in sport departments, ergometric data from clubs' athletes and data from these groups mobile phone. In addition, fans and members of the club participated

5 DataVaults D9.2 Ethics and Data Management Plan.
6 The concept of risk is clarified in Recitals 75-79 of the GDPR.
7 ISACA, "GDPR Data Protection Impact Assessment", 2017

		via questionnaires about their preferences in the sport sector.
2	List of in-scope personal data items	Sport and activity personal data
3	List of key data sources	Olympiacos' CRM platform, Olympiacos Members' mailing list Data from clubs' athletes as files DataVaults Mobile App
4	Limitation in data source: are there limitations that could influence your project's outcomes? Consider: ● bias in data collection, inclusion/exclusion, analysis, algorithms ● gaps or omissions in data ● provenance and data quality ● other issues affecting decisions, such as team composition	Bias in data collection: Maybe a small percentage of fans will not understand the positive values of data collection, especially under the GDPR law and this could lead to data scarcity.
5	Recipients of personal data items. Are you going to be sharing data with other organisations in DataVaults and/or beyond? If so, who? Under what conditions?	Olympiacos did not share any private data without permission of the user. All data is under the control of the user and are being shared only between the user and who he decides to share it with.
6	Period for which the personal data will be stored (e.g., in hours, days, weeks or years, etc.)	Six months on the platform. Olympiacos will not store any data any longer than the time it needs to be exchanged between two pairs. (Olympiacos keeps personal data from every member on CRM platform for one year, until May of every season.)
7	Functional description of the processing operation related to your pilot activities.	Owners will provide data that they want to the DataVaults personal app through its API according to their preferences. The user's data are being anonymised (if the user selects to do so) based on the DataVaults Anonymiser, based on the anonymisation depth levels selected by the user. User Data can be part of anonymised aggregated "personas", if the user selects his data to be included in those Personas. There his data are merged with data from other people and provided as average data values of a "fictional" individual
8	What are your main use cases (please provide a summary and indicate the DataVaults deliverable where details can be retrieved)? Are you replacing another product? You can also simply list them (with one line description) and insert the reference to the DataVaults deliverable where the complete description can be retrieved	Scenario A: Club fans and members Personal Data Marketplace The main goal of this scenario was to collect information related to the fan's activity and the preferences/likes of individuals who are already fans and members of the club. A secondary goal was to collect location data related to the position of the individual. Evidently, such information could help reorganize the marketing plan (new market segmentation, marketing campaigns for specific target groups, finding specific sponsors, etc.), attract new sponsors based on personas (collective profiles) extracted from DataVaults, improve the services offered to the fans and members and finally, increase the revenues and online presence of the club. Scenario B: Athletes Sports and Activity Data Sharing

		The principal type of data to be collected from athletes in the context of the current demonstrator was: • Athletic activity data (e.g., from wearable sensors during practice) • Ergometric and medical examination data (e.g., from club measurements recorded at the beginning of the season) At the level of athletes, Data Seekers could be an athletic equipment companies who are interested in searching through the athletic activity data of athletes of the club, in order to offer better customized and targeted products and services (e.g., depending on what sport each athlete does).
9	DataVaults technological assets (services, tools,) involved in the data collection and/or processing in your pilot activities	• Data Fetcher • Anonymiser • SEAS Analytics Engine
Choice and Consent		
10	How do you provide choices, when appropriate to data subjects regarding use of their personal data? Is consent obtained before using personal information for specific purposes, as required by GDPR? How?	• Initial Consent is asked upon registration to the platform. However, NO data are shared with anyone at that stage. • Data Subjects have the option to delete their data and/or their account at any time by visiting their profile/account page and asking the system to automatically delete all their data (does not apply to data already acquired through a contract by a data seeker) • The Personal App of DataVaults allows users to fully manage their personal data by deleting it and revoking access to it (if not already acquired by a data seeker) • Upon sharing data, Users are provided by the Data Sharing configurator with a dialogue requesting them to verify their sharing options of those data.
11	Are consents (once obtained) appropriately documented and maintained? How?	The agreement is stored in the backend database of the Personal App.
12	What data subject's privacy harms might occur? What risk-mitigation measures have you in place?	In DataVaults the user is in total control regarding the sharing of owned data, while the Risk Management dashboard informs data subjects on their privacy exposure. In addition, prior sharing in the cloud platform of DataVaults, data are anonymised and encrypted; this is one of the benefits of DataVaults towards privacy. Sharing between Olympiacos CRM and DataVaults personal app is also encrypted. Also, as data is only stored temporarily on our backend only small parts of the entire personal record is available at a time so very little amount of data is available for an attacker. It considerably reduces the harms it may cause as only pieces of data are more difficult to interpret. At any time, the sharing can be controlled by the user.
Legitimate Purpose Specification and Use Limitation		
13	Are you going to collect only the personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which the data are processed, in support of	Only relevant data are going to be collected with informed consent procedures, in full compliance with GDPR. We will only collect activity data and other data about their sport activities and preferences, hobbies

	data-minimisation requirements? Is the personal data processing lawful and necessary given the purposes for which the data were collected? Please provide details.	and personal preferences through questionnaires in order to develop the scenarios of the project. In every case, the participant will decide which data want to share with us, giving us permission and will be always informed of what data is shared and with which purpose. The participant will be always in control of the data they share.
Personal Information and Sensitive Information Life Cycle		
14	Are you going to keep personal data for no longer than necessary to support the purposes for which they were collected?	No, we will collect personal data for the project purposes, so only during its deployment.
15	Are you going to collect and/or process one or more of the following types of personal data? If yes, please provide details. a) data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership b) genetic data c) biometric data for the purpose of uniquely identifying a natural person d) data concerning health e) data concerning a natural person's sex life or sexual orientation	No, we are not going to collect any of these types of data
16	How do you ensure that, by default, only personal data which are necessary for each specific processing purpose are processed?	The data subject chooses which data to make available over the platform. The user will also be able to choose what data can be upload on the platform.
Accuracy and Quality		
17	How do you ensure that personal data are kept accurate and up to date, as necessary, and correct personal data errors without delay?	Data are updated following the command of the data subject. Once the connection to the platform is done, all updates are automatically forwarded to the platform (unless the user decides to stop the sharing).
Openness, Transparency and Notice		
18	Are you going to process personal data fairly, transparently and in compliance with applicable legal requirements?	Yes
19	How do you communicate to data subjects their rights, notices, and answer their questions and provide information to them relating to data processing, in a manner that is clear, easy to understand, and age appropriate to the data subject? Are you going to provide the data subjects, at the time personal data are obtained from them, all necessary information elements? How do you restrict use of their associated personal data?	By default, NO personal data is shared, so this question does not apply. However, once data is shared, we cannot control this, as this is something that falls out of the scope of the project. Maybe we could have some smart contracts which could say something “generic” about to no harm, etc. but again this is not controllable by the system. Regarding the demonstrators, we said that for each demonstrator, its users will have a look and accept the terms of that specific demonstrators. But nothing more is expected. All relevant information will be provided at the sign-up /registration page, and in the relevant terms. A FAQ page will be delivered to answer most common questions and a functional mailbox will be provided for further questions.
Individual Participation		
20	How are the data subjects allowed to withdraw consent to use their associated	By directly removing their data and/or their account. In addition, they also must disable sharing of data to

	personal data at any time, as long as the withdrawal does not result in legal violations about which you have informed the data subjects?	ensure no data stays available but is never retrieved by the DataVaults platform.
21	How do you ensure that a data subject can exercise his or her rights mandated by the GDPR? And do you have processes to provide documented reasons for denying requests?	Data Subjects can at any time delete their account and remove their personal data. No requests will be denied. Olympiacos already provides procedures to exercise their rights against GDPR.
22	How do you allow the data subject to obtain confirmation regarding whether or not personal data are being processed (including personal data used in partnership with other controllers), and when that is the case, to provide the data subject access to the associated personal data and related mandatory information (including, for instance, information concerning the purposes, the data categories, the recipients, the retention periods, the rights for deletion and registering complaints, etc.) in compliance with GDPR requirements?	Such information is stored in the blockchain where the data shared can be viewed by the data subject. This is provided as part of a dashboard for each user, where he can see: • The data he has already collected but he holds without sharing them. • The data he has already shared and under which terms. (E.g., anonymised or not, for how long, etc.) • The data that (after being shared) have been acquired by a data seeker and the relevant transaction ID. Data processing involved in the Olympiacos demonstration, and their purposes are specified in the consent and information sheet related to the demonstration.
23	Are you able to provide to the data subject, upon request and without prejudice, a copy of his/her personal data? How? Can this copy be delivered in a commonly used digital format and for free or for a reasonable fee (where the fee is based on actual administrative costs)?	Yes. Data Subjects can download a copy of the data they have provided to the platform for free.
Accountability		
24	Have you established an appropriately qualified data protection officer (if applicable, according to the GDPR)? Can the data subjects contact him/her for any issue related to processing of their personal data or to the exercise of their rights under GDPR?	The Olympiacos' lawyer has been appointed and is available as imposed by the national and EU laws.
25	Which DataVaults tool in your demonstrator support accountability?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access
Security Safeguards		
26	How do you ensure that appropriate technical and/or organisational safeguards are implemented to secure personal data, including protections against unauthorized or unlawful processing and against accidental loss, destruction or damage? What DataVaults technologies/tools ensure a level of security for the personal data appropriate to the personal harm risk, including, as appropriate: a) pseudonymisation and/or encryption. b) procedures to establish confidentiality, integrity, availability and resilience of processing systems and services, data backup and recovery; and	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access • Anonymisation Engine • Data Encryption Engine

	c) Regular testing of associated security controls?	
27	How do you assess the privacy risks and/or the likelihood of privacy harms to data subjects, for instance in the event of unauthorised access, sharing or use of personal data and/or unauthorised or accidental destruction, loss, or alteration of personal data?	Risk Management Dashboard can assess the privacy risks for the shared datasets of the user, based on the sharing policies applied while data are protected from unauthorised access within the platform. Also, with the Blockchain infrastructure for storing transactions, we ensure that no unnoticed alternation, deletion or access will be performed.
Monitoring, Measuring and Reporting		
28	Do you provide reports for data subjects (at specified times; upon their request as appropriate; and reflecting all components required by GDPR), including: a) personal data-breach notifications. b) reports regarding correction of personal data erasure and/or incorrect personal data. c) reports showing the content of personal data associated to the subject. d) reports showing the personal data associated to the data subject shared with others, including the reasons for such sharing; and e) Full digital copies of personal data transmitted directly to another data controller in support of data portability requirements.	• Dashboard of collected data. • Transactions Dashboard
29	How do you maintain a record of processing activities involving personal data that includes a) the name and contact details for your enterprise/institution; b) the purposes for processing personal data; c) a description of the categories of data subjects and of the categories of personal data involved in the processing; d) the categories of recipients to whom the personal data have been or will be disclosed; and where possible, the established time limits for erasure of the different categories of data?	There is a record showing which data seeker has acquired access to which data set, which is retrieved via the blockchain system and displayed to the user via his sharing dashboard page. Also, information about the prospective categories of data seekers that can have access on each shared dataset is provided in the sharing configuration of each shared dataset, in the access policies lists defined by the data subject.
Preventing Harm		
30	Do you allow the data subjects to request removal of their personal data from automated processing and profiling in situations that could result in adverse legal effects or harms to them?	Yes, in case their data is not already acquired by a third party (as then any data processing happens off the platform and cannot be controlled). See the first two questions under “Individual Participation”
31	How do you ensure that data subjects who exercise their rights (for changing how their personal data are used; request copies of personal data; and/or exercise other rights under GDPR) do not adversely affect the rights and freedoms of others or do not infringe a contract or a binding provision?	The exercise of rights concerns only data not already acquired by another party
Third-party/Vendor Management		
32	Do you have documented third-party / vendor management policies (and supporting procedures) to specify the type of documented contract (in hard copy and/or	As regards the smart contracts, our policies rely on the DataVaults sharing configuration tool to set all the specific characteristics and features of the data sharing (including, for instance, the selection of the

	digital form), or other legal act under union or member-state law, setting out key aspects, like the subject-matter and duration of the processing, the nature and purpose of the processing and the types of personal data and the categories of data subjects?	privacy/anonymisation level, of the visibility level, of the access policies level, of the pricing and of the licensing). The Individual can share not only her/his personal data collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase. In this way, whenever an Individual decides to make data available to Data Seekers through the DataVaults platform, they are prompted to define various details regarding data sharing, enabling the construction of fine-grained policies. The various parameters of a data asset sharing configuration can be modified at a later stage by the Individual, who can also completely revoke access to the shared data asset, by appropriately configuring the bound access policies. For the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period.
Breach Management		
33	Do you have documented personal data breach policies and/or tools for: a) notifying appropriate supervisory authorities of the breach in a timely manner, and with reasons provided for any delays. b) notifying data subjects of high-risk breaches (as defined by GDPR) no later than 72 hours after discovery of a breach, if it is determined (following documented procedures for performing harm risk analysis) that the personal data breach will result in privacy harm to the associated data subjects; and c) Including all items necessary, within the notice, as required by GDPR?	Olympiacos does not have any specific documentation related to data breach policy or such tools but falls under its GDPR compliant policy. Tools are already in place if needed.
Security and Privacy by Design		
34	Have the security and privacy protections been embedded into the full lifecycle of automated decision-making processes involving personal data to safeguard the data subject's rights, freedoms and legitimate interests? Have you implemented appropriate technical security and privacy controls, supported by documented privacy principles (e.g., the ISACA Privacy Principles, and/or IEEE privacy standards, etc.), in order to appropriately mitigate harms to individuals to the extent possible in compliance with GDPR and to protect the rights of data subjects?	Yes, this is the main aim of the DataVaults project. Data subjects are able to on-board the platform and control which data they would like to take on-board, then decide if they want to share them, with which data seeker categories, under which provisions, and remove them if they want to (unless they have been engaged in a binding contract) Core to these processes are the following components: Data Sharing Configurator; Anonymiser; Access Policy Editor; Data Sharing and Confirmation Dashboard; Risk Management Dashboard; Blockchain Infrastructure; Smart Sharing Contracts Olympiacos also already provide privacy protections by design in its lifecycle and fully respect GDPR.
Legal, Ethical and Societal Implications		
35	Do people understand your purpose – especially people who the data are about or who are impacted by their use? How have you been communicating your purpose? Has this communication been clear?	The user that wants to share its data to the DataVaults platform is first informed of all the consequences through an information sheet and an informed consent to sign before being able to configure any exchange of data. We try to present it in the most user-friendly way

	How are you ensuring more vulnerable individuals or groups understand?	on the smartphone to ensure that it will be correctly read and understood by the majority. During the demonstration activities, participants will be selected according to certain criteria. If more vulnerable individuals are selected, we will ensure that a responsible person can help them during this process.
36	What existing ethical codes apply to your sector or project? Are there any legislation, policies, or other regulation shape how you use data, in addition to those listed in DataVaults D2.3? Are you in compliance with them (if applicable)?	GDPR regulation.
37	Positive effects on people. Which individuals, groups, demographics or organisations will be positively affected by your DataVaults pilot? How? How are you measuring and communicating positive impact? How could you increase it?	Our main objectives are improving the app to help users to take care of their activities and ease collection of data for companies that build a better management-marketing system for the future. The KPIs we are planning to measure should help have a sense of those positive impacts. The analysis of those as well as feedback received during the demonstration will help to improve any positive aspect. We are also communicating regularly and widely on our various social media about the positive impacts of our work.
38	Negative effects on people. Who could be negatively affected by your DataVaults pilot? How? Could the way that data is collected, used or shared cause any harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access to some service? How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.	We do not foresee any negative impact on people, related to the three scenarios and the type of data collected.
39	Openness and transparency in the design process How open can you be about this project/DataVaults pilot? Are you asking the volunteer for feedback on the project and its outcomes? Are you building in thoughts, ideas and considerations of people affected by your project/DataVaults pilot? How? What information or training might be needed to help people understand data issues?	Any detail about the project will be communicated in the most transparent way at any time of the project. Feedbacks will be collected on a voluntary basis but also according to a predefined time frame in order to get as much feedback we can get to help us improve our communication and our services. Any feedback will be handled individually and in a transparent way.
40	Review and iterations Are you planning to measure, monitor and discuss the data ethics issues in the post-project phase and to review this EDPIA?	No plan is in place for measuring and monitoring data ethics issues in the post-project phase.

2.3.1.2 *Demonstrator #2 – Strengthening Entrepreneurship and Mobility*

Introductory questions		
		Responses
1	Nature, scope, context and purposes of the processing (Recital 90)	Mobility, cultural, commercial preferences information collected within the DataVaults app to be used in order to provide services to citizens, tourists and visitors
2	List of in-scope personal data items	• Mobility information (position, route) • Commercial preferences (where, how and what are people buying) • Cultural and touristic preferences (museum, music, food, entertainment preferences)
3	List of key data sources	DataVaults app
4	Limitation in data source: are there limitations that could influence your project's outcomes? Consider: ● bias in data collection, inclusion/exclusion, analysis, algorithms ● gaps or omissions in data ● provenance and data quality ● other issues affecting decisions, such as team composition	Limited availability of data might be an issue. Collection of GPS information regarding traffic routes and parking spots is challenging.
5	Recipients of personal data items. Are you going to be sharing data with other organisations in DataVaults and/or beyond? If so, who? Under what conditions?	The Municipality is the sole recipient of personal data. No sharing of personal data with others is foreseen. Only results will be shared.
6	Period for which the personal data will be stored (e.g., in hours, days, weeks or years, etc.)	6 months.
7	Functional description of the processing operation related to your pilot activities.	The user's data are being anonymised (if the user selects to do so) based on the DataVaults Anonymiser, based on the anonymisation depth levels selected by the user. User Data can be part of anonymised aggregated "personas", if the user selects his data to be included in those Personas. There his data are merged with data from other people and provided as average data values of a "fictional" individual
8	What are your main use cases (please provide a summary and indicate the DataVaults deliverable where details can be retrieved)? Are you replacing another product? You can also simply list them (with one line description) and insert the reference to the DataVaults deliverable where the complete description can be retrieved	Scenario A - Smart Mobility Services for Individuals. Goal: To better schedule the mobility strategy and the relevant services within the city of Piraeus around sport venues at game time Scenario B - Empowering local entrepreneurship. Goal: to strengthen the local economy through activities that can be brought forward by the municipality, based on consumer behaviours and preferences Scenario C - Services for Personalized cultural and touristic experiences. Goal: Create services that target tourists and visitors in the city of Piraeus, based their individual interests
9	DataVaults technological assets (services, tools…) involved in the data collection and/or processing in your pilot activities.	• Data Fetcher • Edge Analytics Engine • Anonymiser

		• SEAS Analytics Engine
Choice and Consent		
10	How do you provide choices, when appropriate to data subjects regarding use of their personal data? Is consent obtained before using personal information for specific purposes, as required by GDPR? How?	• Initial Consent is asked upon registration to the platform. However, NO data are shared with anyone at that stage. • Data Subjects have the option to delete their data and/or their account at any time by visiting their profile/account page and asking the system to automatically delete all their data. (Does not apply to data already acquired through a contract by a data seeker) • The Personal App of DataVaults allows users to fully manage their personal data by deleting it and revoking access to it (if not already acquired by a data seeker) • Upon sharing data, users are provided by the Data Sharing configurator with a dialogue requesting them to verify their sharing options of those data.
11	Are consents (once obtained) appropriately documented and maintained? How?	The agreement is stored in the backend database of the Personal App.
12	What data subject's privacy harms might occur? What risk-mitigation measures have you in place?	In DataVaults the user is in total control regarding the sharing of owned data, while the Risk Management dashboard informs data subjects on their privacy exposure. In addition, prior sharing in the cloud platform of DataVaults, data are anonymised and encrypted; this is one of the benefits of DataVaults towards privacy.
Legitimate Purpose Specification and Use Limitation		
13	Are you going to collect only the personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which the data are processed, in support of data-minimization requirements? Is the personal data processing lawful and necessary given the purposes for which the data were collected? Please provide details.	Only relevant data are going to be collected with informed consent procedures, in full compliance with GDPR. The collection is therefore lawful.
Personal Information and Sensitive Information Life Cycle		
14	Are you going to keep personal data for no longer than necessary to support the purposes for which they were collected?	Yes, we store it only for the scope of the project
15	Are you going to collect and/or process one or more of the following types of personal data? If yes, please provide details. a) data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership b) genetic data c) biometric data for the purpose of uniquely identifying a natural person d) data concerning health e) data concerning a natural person's sex life or sexual orientation	No to all
16	How do you ensure that, by default, only personal data which are necessary for each	The data subject chooses which data to make available over the platform.

	specific processing purpose are actually processed?	
Accuracy and Quality		
17	How do you ensure that personal data are kept accurate and up to date, as necessary, and correct personal data errors without delay?	Data are updated following the command of the data subject.
Openness, Transparency and Notice		
18	Are you going to process personal data fairly, transparently and in compliance with applicable legal requirements?	Yes
19	How do you communicate to data subjects their rights, notices, and answer their questions and provide information to them relating to data processing, in a manner that is clear, easy to understand, and age appropriate to the data subject? Are you going to provide the data subjects, at the time personal data are obtained from them, all necessary information elements? How do you restrict use of their associated personal data?	By default, NO personal data is shared, so this question does not apply. However, once data is shared, we cannot control this, as this is something that falls out of the scope of the project. Maybe we could have some smart contracts which could say something “generic” about to no harm, etc. but again this is not controllable by the system. All relevant information will be provided at the sign-up /registration page, in the relevant terms. An FAQ page will be delivered to answer most common questions and a functional mailbox will be provided for further questions.
Individual Participation		
20	How are the data subjects allowed to withdraw consent to use their associated personal data at any time, as long as the withdrawal does not result in legal violations about which you have informed the data subjects?	By directly removing their data and/or their account.
21	How do you ensure that a data subject can exercise his or her rights mandated by the GDPR? And do you have processes to provide documented reasons for denying requests?	Data Subjects can at any time delete their account and remove their personal data. No requests will be denied. Municipality of Piraeus has as main contact point its DPO (dpo@piraeus.gov,gr) but is also obliged by law to allow people exercise their rights against GDPR. Any request made through any channel available (telephone, email, official document or other) is received and forwarded to the responsible Directorate within the Municipality. A formal answer to any request must be provided within the timeframe set by Greek law.
22	How do you allow the data subject to obtain confirmation regarding whether or not personal data are being processed (including personal data used in partnership with other controllers), and when that is the case, to provide the data subject access to the associated personal data and related mandatory information (including, for instance, information concerning the purposes, the data categories, the recipients, the retention periods, the rights for deletion and registering complaints, etc.) in compliance with GDPR requirements?	Such information is stored in the blockchain where the data shared can be viewed by the data subject. This is provided as part of a dashboard for each user, where he can see: • the data he has already collected but he holds without sharing them. • The data he has already shared and under which terms (e.g., anonymised or not, for how long, etc.) • The data that (after sharing them) have been acquired by a data seeker and the relevant transaction id.

23	Are you able to provide to the data subject, upon request and without prejudice, a copy of his/her personal data? How? Can this copy be delivered in a commonly used digital format and for free or for a reasonable fee (where the fee is based on actual administrative costs)?	Yes. Data Subjects can download a copy of the data they have provided to the platform for free
Accountability		
24	Have you established an appropriately qualified data protection officer (if applicable, according to the GDPR)? Can the data subjects contact him/her for any issue related to processing of their personal data or to the exercise of their rights under GDPR?	The Municipality of Piraeus DPO has been appointed and is available as imposed by the national and EU laws.
25	Which DataVaults tool in your demonstrator support accountability?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access
Security Safeguards		
26	How do you ensure that appropriate technical and/or organisational safeguards are implemented to secure personal data, including protections against unauthorized or unlawful processing and against accidental loss, destruction or damage? What DataVaults technologies/tools ensure a level of security for the personal data appropriate to the personal harm risk, including, as appropriate: a) pseudonymisation and/or encryption. b) procedures to establish confidentiality, integrity, availability and resilience of processing systems and services, data backup and recovery; and c) Regular testing of associated security controls?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access • Anonymisation Engine • Data Encryption Engine • TPM Attestation for devices running DataVaults
27	How do you assess the privacy risks and/or the likelihood of privacy harms to data subjects, for instance in the event of unauthorised access, sharing or use of personal data and/or unauthorised or accidental destruction, loss, or alteration of personal data?	Risk Management Dashboard can assess the privacy risks for the shared datasets of the user, based on the sharing policies applied while data are protected by unauthorised access within the platform. Also, with the blockchain infrastructure for storing transactions, we ensure that no unnoticed alternation, deletion or access will be performed.
Monitoring, Measuring and Reporting		
28	Do you provide reports for data subjects (at specified times; upon their request as appropriate; and reflecting all components required by GDPR), including: a) personal data-breach notifications. b) reports regarding correction of personal data erasure and/or incorrect personal data. c) reports showing the content of personal data associated to the subject. d) reports showing the personal data associated to the data subject shared with others, including the reasons for such sharing; and	a) These are provided upon requests via their Inbox page. b) These are provided upon requests via their Inbox page. c) Dashboard of collected data. d) Transactions Dashboard. e) Transactions Dashboard.

	e) Full digital copies of personal data transmitted directly to another data controller in support of data portability requirements.	
29	How do you maintain a record of processing activities involving personal data that includes a) the name and contact details for your enterprise/institution; b) the purposes for processing personal data; c) a description of the categories of data subjects and of the categories of personal data involved in the processing; d) the categories of recipients to whom the personal data have been or will be disclosed; and where possible, the established time limits for erasure of the different categories of data?	There is a record showing which data seeker has acquired access to which data set, which is retrieved via the blockchain system and displayed to the user via his sharing dashboard page. Also, information about the prospective categories of data seekers that can have access on each shared dataset is provided in the sharing configuration of each shared dataset, in the access policies lists defined by the data subject.
Preventing Harm		
30	Do you allow the data subjects to request removal of their personal data from automated processing and profiling in situations that could result in adverse legal effects or harms to them?	Yes, in case their data is not already acquired by a third party (as then any data processing happens off the platform and cannot be controlled). See the first two questions under “Individual Participation”.
31	How do you ensure that data subjects who exercise their rights (for changing how their personal data are used; request copies of personal data; and/or exercise other rights under GDPR) do not adversely affect the rights and freedoms of others or do not infringe a contract or a binding provision?	The exercise of rights concerns only data not already acquired by another party.
Third-party/Vendor Management		
32	Do you have documented third-party / vendor management policies (and supporting procedures) to specify the type of documented contract (in hard copy and/or digital form), or other legal act under union or member-state law, setting out key aspects, like the subject-matter and duration of the processing, the nature and purpose of the processing and the types of personal data and the categories of data subjects?	As regards the smart contracts, our policies rely on the DataVaults sharing configuration tool to set all the specific characteristics and features of the data sharing (including, for instance, the selection of the privacy/anonymisation level, of the visibility level, of the access policies level, of the pricing and of the licensing). The Individual can share not only their personal data collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase. In this way, whenever an Individual decides to make data available to Data Seekers through the DataVaults platform, they are prompted to define various details regarding data sharing, enabling the construction of fine-grained policies. The various parameters of a data asset sharing configuration can be modified at a later stage by the Individual, who can also completely revoke access to the shared data asset, by appropriately configuring the bound access policies. For the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period.
Breach Management		
33	Do you have documented personal data breach policies and/or tools for:	The policies were developed for the v0.5 and v1.0 of the platform, so after v0.5 we will support the

	a) notifying appropriate supervisory authorities of the breach in a timely manner, and with reasons provided for any delays. b) notifying data subjects of high-risk breaches (as defined by GDPR) no later than 72 hours after discovery of a breach, if it is determined (following documented procedures for performing harm risk analysis) that the personal data breach will result in privacy harm to the associated data subjects; and c) Including all items necessary within the notice as required by GDPR?	notification of the data subjects for the case of breaches.
Security and Privacy by Design		
34	Have the security and privacy protections been embedded into the full lifecycle of automated decision-making processes involving personal data to safeguard the data subject's rights, freedoms and legitimate interests? Have you implemented appropriate technical security and privacy controls, supported by documented privacy principles (e.g., the ISACA Privacy Principles, and/or IEEE privacy standards, etc.), in order to appropriately mitigate harms to individuals to the extent possible in compliance with GDPR and to protect the rights of data subjects?	Yes, this is the main aim of the DataVaults project. Data subjects are able to on-board the platform and control which data they would like to take on-board, then decide if they want to share them, with which data seeker categories, under which provisions, and remove them if they want to. (Unless they have been engaged in a binding contract) Core to these processes are the following components: Data Sharing Configurator; Anonymiser; Access Policy Editor; Data Sharing and Confirmation Dashboard; Risk Management Dashboard; Blockchain Infrastructure; Smart Sharing Contracts.
Legal, Ethical and Societal Implications		
35	Do people understand your purpose – especially people who's the data are about or who are impacted by their use? How have you been communicating your purpose? Has this communication been clear? How are you ensuring more vulnerable individuals or groups understand?	The user that wants to share their data to the DataVaults platform is first informed of all the consequences through an information sheet and an informed consent to sign before being able to configure any exchange of data. We try to present it in the most user-friendly way on the smartphone to ensure that it will be correctly read and understood by the majority. During the demonstration activities, participants will be selected according to certain criteria. If more vulnerable individuals are selected, we will ensure that a responsible person can help them during this process.
36	What existing ethical codes apply to your sector or project? Are there any legislation, policies, or other regulation shape how you use data, in addition to those listed in DataVaults D2.3? Are you in compliance with them (if applicable)?	Only GDRP applies.
37	Positive effects on people. Which individuals, groups, demographics or organisations will be positively affected by your DataVaults pilot? How? How are you measuring and communicating positive impact? How could you increase it?	All scenarios are either targeting the improvement of the life of citizens and tourists (mobility and tailored touristic/cultural experiences) or the enhancement of the local economy (empowering local shop owners and entrepreneurs). We are planning surveys in order to measure the expected positive impact, as described in the DataVaults Action Plan.

38	Negative effects on people. Who could be negatively affected by your DataVaults pilot? How? Could the way that data is collected, used or shared cause any harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access to some service? How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.	We do not foresee any negative impact on people, related to the three scenarios and the type of data collected.
39	Openness and transparency in the design process How open can you be about this project/DataVaults pilot? Are you asking the volunteer for feedback on the project and its outcomes? Are you building in thoughts, ideas and considerations of people affected by your project/ DataVaults pilot? How? What information or training might be needed to help people understand data issues?	As a local government authority, we are open about the project, if not in contrast with national and EU laws. A small group of volunteers are asked for feedback, in order to help pass on ideas regarding the project. Ideas and thoughts on training material and people affected will be considered after finalizing the pilots and testing the DataVaults app.
40	Review and iterations Are you planning to measure, monitor and discuss the data ethics issues in the post-project phase and to review this EDPIA?	No plan is in place for measuring and monitoring data ethics issues in the post-project phase.

2.3.1.3 *Demonstrator #3 – Healthcare Data Retention and Sharing*

Introductory questions		
		Responses
1	Nature, scope, context and purposes of the processing. (Recital 90)	Personal health and medical data collected in a personal mobile app to give patients control of their health data (gather all health data about them and their family). Ease running clinical trials. (Discover patients, collect answers to questionnaires, …)
2	List of in-scope personal data items	Health and medical data. Those can be in the context of a clinical trial (fake during demonstration activities)
3	List of key data sources	Andaman7 application that can gather data from other sources. Any connected device that can provide health/medical data (will be tested as a proof of work during demonstration).
4	Limitation in data source: are there limitations that could influence your project's outcomes? Consider: ● bias in data collection, inclusion/exclusion, analysis, algorithms ● gaps or omissions in data	Data analysis and calculation algorithms should be done with great care and provided to the user with explanations to prevent any bad understanding of such data. Completeness of data is always a challenge as the health situation of a person evolves all the time. So, this is more of a "best effort" nature. Producers of data (doctors, hospitals etc.) are sometimes reluctant to

	● provenance and data quality ● other issues affecting decisions, such as team composition	share the data they generate about patients, sometimes for good (but not always) reasons. All data should contain the exact source of data to be able to identify data quality and provenance. The user will be responsible for data exclusion depending on its quality.
5	Recipients of personal data items. Are you going to be sharing data with other organisations in DataVaults and/or beyond? If so, who? Under what conditions?	All data is under the control of the user and will be shared only between the user and who he is willing to share with. In the special use case of clinical trials, data will be shared with the care facility/facilities that run the clinical trials. The user will be informed and will sign a consent before entering the trial. In any other use case, data will only be shared between Andaman7 and the DataVaults app (under user consent) and any extra exchange will be configured by the user through the interface of the DataVaults platform (under user consent).
6	Period for which the personal data will be stored (e.g., in hours, days, weeks or years, etc.)	Six months on the DataVaults platform. Andaman7 will not store any data any longer than the time it needs to be exchanged between two pairs.
7	Functional description of the processing operation related to your pilot activities	The user's data will be transferred from its Andaman7 application to the DataVaults personal app through its API according to their preferences. Only exchange and storage are involved in this phase. The user's data are being anonymised (if the user selects to do so) based on the DataVaults Anonymiser, based on the anonymisation depth levels selected by the user. User Data can be part of anonymised aggregated "personas", if the user selects his data to be included in those Personas. There his data are merged with data from other people and provided as average data values of a "fictional" individual.
8	What are your main use cases (please provide a summary and indicate the DataVaults deliverable where details can be retrieved)? Are you replacing another product? You can also simply list them (with one line description) and insert the reference to the DataVaults deliverable where the complete description can be retrieved	Main use cases are: ● Sharing of Andaman7 collected data to the DataVaults platform to backup data and allow external exchange configuration. ● Running clinical trials with data storage on the DataVaults platform. ● Collecting medical/health data from new sources through the DataVaults platform. ● Aggregation of some basic health data to improve content of the health record. Details can be found in deliverable 1.4
9	DataVaults technological assets (services, tools…) involved in the data collection and/or processing in your pilot activities	● Data Fetcher ● Anonymiser ● SEAS Analytics Engine
Choice and Consent		
10	How do you provide choices, when appropriate, to data subjects regarding use of their personal data? Is consent obtained before using personal information for specific purposes, as required by GDPR? How?	Data transferred from Andaman7 to the DataVaults personal app is configurable by the citizen. During the onboarding process, the citizen can select which part of the medical data he/she wants to share to the DataVaults platform.

		Initial consent is asked upon registration to the platform. However, NO data is shared with anyone at that stage. Data Subjects have the option to delete their data and/or their account at any time by visiting their profile/account page and asking the system to automatically delete all their data. (Does not apply to data already acquired through a contract by a data seeker) The Personal App of DataVaults allows users to fully manage their personal data by deleting it and revoking access to it. (If not already acquired by a data seeker) Upon sharing data, Users are provided by the Data Sharing configurator with a dialogue requesting them to verify their sharing options of those data.
11	Are consents (once obtained) appropriately documented and maintained? How?	The agreement is stored in the backend database of the Personal App. Andaman7 also stores consent related to specific activities in its backend database.
12	What data subject's privacy harms might occur? What risk-mitigation measures have you in place?	In DataVaults the user is in total control regarding the sharing of owned data, while the Risk Management dashboard informs data subjects on their privacy exposure. In addition, prior sharing in the cloud platform of DataVaults, data is anonymised and encrypted; this is one of the benefits of DataVaults towards privacy. Sharing between Andaman7 and DataVaults personal app is also encrypted. Also, as data is only stored temporarily on our backend, only small parts of the entire medical record is available at a time so very little amount of data is available for an attacker. It considerably reduces the harm it may cause as just pieces of data are more difficult to interpret. At any time, the sharing can be controlled by the user.
Legitimate Purpose Specification and Use Limitation		
13	Are you going to collect only the personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which the data are processed, in support of data-minimization requirements? Is the personal data processing lawful and necessary given the purposes for which the data were collected? Please provide details.	Only personal medical and health data related to users asking for connection will be collected. The user will be able to configure what data he is willing to share. Processing is limited to storing, forwarding and basic calculation on health data. (Only less sensitive health data such as activities, sleep etc.) At any time, the user can check what data is collected. It is therefore lawful.
Personal Information and Sensitive Information Life Cycle		
14	Are you going to keep personal data for no longer than necessary to support the purposes for which they were collected?	Data will be stored only on the smartphone of the user and on the Personal DataVaults App. Andaman7 server won't keep data any longer than the time it needs to be transferred to its destination. Data stored in the mobile app and in DataVaults will be under the control of the user / citizen. They can delete all the data at any time.
15	Are you going to collect and/or process one or more of the following types of personal data? If yes, please provide details. a) Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership.	Data from those five categories may indeed be part of a health record. It will be collected for the purpose of providing it to the concerned patient or to a healthcare facility (study trials), in all cases after patient consent.

	b) Genetic data. c) Biometric data for the purpose of uniquely identifying a natural person. d) Data concerning health. e) Data concerning a natural person's sex life or sexual orientation.	
16	How do you ensure that, by default, only personal data which are necessary for each specific processing purpose are actually processed?	The data subject chooses which data to make available over the platform. As mentioned earlier, the user is also able to choose what data can be exchanged between Andaman7 and the platform.
Accuracy and Quality		
17	How do you ensure that personal data are kept accurate and up to date, as necessary, and correct personal data errors without delay?	Data is updated following the command of the data subject. Medical data can come from various sources in Andaman7. Most are trusted platforms (hospitals for example) and provide accurate and up-to-date data. For less trusted sources, it's under the responsibility of the data owner to update its data accordingly. Once the connection to the platform is done, all updates are automatically forwarded to the platform (unless the user decides to stop the sharing).
Openness, Transparency and Notice		
18	Are you going to process personal data fairly, transparently and in compliance with applicable legal requirements?	Yes.
19	How do you communicate to data subjects their rights, notices, and answer their questions and provide information to them relating to data processing, in a manner that is clear, easy to understand, and age appropriate to the data subject? Are you going to provide the data subjects, at the time personal data are obtained from them, all necessary information elements? How do you restrict use of their associated personal data?	By default, NO personal data is shared, so this question does not apply. However, once data is shared, we cannot control this, as this is something that falls out of the scope of the project. Maybe we could have some smart contracts which could say something "generic" about to no harm, etc., but again this is not controllable by the system. Regarding the demonstrators, we said that for each demonstrator, its users will have a look and accept the terms of that specific demonstrators. But nothing more is expected. All relevant information will be provided at the sign-up /registration page, in the relevant terms. An FAQ page will be delivered to answer most common questions and a functional mailbox will be provided for further questions.
Individual Participation		
20	How are the data subjects allowed to withdraw consent to use their associated personal data at any time, as long as the withdrawal does not result in legal violations about which you have informed the data subjects?	By directly removing their data and/or their account. In addition, they also must disable sharing of data from the Andaman7 to ensure no data stays available but is never retrieved by the DataVaults platform.
21	How do you ensure that a data subject can exercise his or her rights mandated by the GDPR? And do you have processes to provide documented reasons for denying requests?	Data Subjects can at any time delete their account and remove their personal data. No requests will be denied. Andaman7 already provides procedures to exercise his or her rights against GDPR. They can simply contact us via support@andaman7.com and we will answer as best as possible to their request.
22	How do you allow the data subject to obtain confirmation regarding whether or not	Such information is stored in the blockchain where the data shared can be viewed by the data subject.

	personal data are being processed (including personal data used in partnership with other controllers), and when that is the case, to provide the data subject access to the associated personal data and related mandatory information (including, for instance, information concerning the purposes, the data categories, the recipients, the retention periods, the rights for deletion and registering complaints, etc.) in compliance with GDPR requirements?	This is provided as part of a dashboard for each user, where he can see: ● The data he has already collected but he holds without sharing them. ● And the data he has already shared and under which terms (e.g., anonymised or not, for how long, etc.) ● The data that (after sharing them) have been acquired by a data seeker and the relevant transaction id. Data processing involved in the Andaman7 demonstration, and their purposes are specified in the consent and information sheet related to the demonstration.
23	Are you able to provide to the data subject, upon request and without prejudice, a copy of his/her personal data? How? Can this copy be delivered in a commonly used digital format and for free or for a reasonable fee (where the fee is based on actual administrative costs)?	Yes. Data Subjects can download a copy of the data they have provided to the platform for free. Andaman7 can also provide a copy of data they have upon email request.
Accountability		
24	Have you established an appropriately qualified data protection officer (if applicable, according to the GDPR)? Can the data subjects contact him/her for any issue related to processing of their personal data or to the exercise of their rights under GDPR?	Any question or request addressed to support@andaman7.com, will be processed by a member of our support team and transferred to the most appropriate member of the team. Our DPO can be directly contacted via dpo@andaman7.com.
25	Which DataVaults tool in your demonstrator support accountability?	Blockchain infrastructure for storing transactions. Access Policy Engine for safekeeping access.
Security Safeguards		
26	How do you ensure that appropriate technical and/or organisational safeguards are implemented to secure personal data, including protections against unauthorized or unlawful processing and against accidental loss, destruction or damage? What DataVaults technologies/tools ensure a level of security for the personal data appropriate to the personal harm risk, including, as appropriate: a) pseudonymisation and/or encryption. b) procedures to establish confidentiality, integrity, availability and resilience of processing systems and services, data backup and recovery; and c) Regular testing of associated security controls?	Blockchain infrastructure for storing transactions. Access Policy Engine for safekeeping access. Anonymisation Engine. Data Encryption Engine. TPM Attestation for devices running DataVaults. A copy of all data is kept locally on the Andaman7 application.
27	How do you assess the privacy risks and/or the likelihood of privacy harms to data subjects, for instance in the event of unauthorised access, sharing or use of personal data and/or unauthorised or accidental destruction, loss, or alteration of personal data?	Risk Management Dashboard can assess the privacy risks for the shared datasets of the user, based on the sharing policies applied while data are protected by unauthorised access within the platform. Also, with the Blockchain infrastructure for storing transactions, we ensure that no unnoticed alternation, deletion or access will be performed. Andaman7 doesn't allow any data modification on the smartphone, only addition of data. History is always

		available. Any exchange of data is encrypted and therefore protected. (Altering encrypted data will lead to unreadable data)
Monitoring, Measuring and Reporting		
28	Do you provide reports for data subjects (at specified times; upon their request as appropriate; and reflecting all components required by GDPR), including: a) personal data-breach notifications. b) reports regarding correction of personal data erasure and/or incorrect personal data. c) reports showing the content of personal data associated to the subject. d) reports showing the personal data associated to the data subject shared with others, including the reasons for such sharing; and e) Full digital copies of personal data transmitted directly to another data controller in support of data portability requirements.	a) These are provided upon requests via their Inbox page. Andaman7 will report any breach to their system by email as soon as it is detected. b) These are provided upon requests via their Inbox page or via email. c) Dashboard of collected data. Andaman7 can provide data reports upon request but almost no personal data is stored on Andaman7 backend. Medical data is only available on their smartphone. d) Transactions Dashboard and Andaman7 app. e) Transactions Dashboard. Andaman7 cannot provide a full copy of data transmitted as we don't store any data. We can only provide some information about the time lapse the sharing occurs. It's up to the user to associate it with data stored on their smartphone.
29	How do you maintain a record of processing activities involving personal data that includes a) the name and contact details for your enterprise/institution; b) the purposes for processing personal data; c) a description of the categories of data subjects and of the categories of personal data involved in the processing; d) the categories of recipients to whom the personal data have been or will be disclosed; and where possible, the established time limits for erasure of the different categories of data?	There is a record showing which data seeker has acquired access to which data set, which is retrieved via the blockchain system and displayed to the user via his sharing dashboard page. Also, information about the prospective categories of data seekers that can have access on each shared dataset is provided in the sharing configuration of each shared dataset, in the access policies lists defined by the data subject. Additionally, Andaman7 backend database contains some information about the sharing that occurs on the smartphone. Such information can be easily associated with a specific purpose, destination and category of recipient.
Preventing Harm		
30	Do you allow the data subjects to request removal of their personal data from automated processing and profiling in situations that could result in adverse legal effects or harms to them?	Yes, in case their data is not already acquired by a third party (as then any data processing happens off the platform and cannot be controlled). See the first two questions under "Individual Participation".
31	How do you ensure that data subjects who exercise their rights (for changing how their personal data are used; request copies of personal data; and/or exercise other rights under GDPR) do not adversely affect the rights and freedoms of others or do not infringe a contract or a binding provision?	The exercise of rights concerns only data not already acquired by another party.
Third-party/Vendor Management		
32	Do you have documented third-party / vendor management policies (and supporting procedures) to specify the type of documented contract (in hard copy and/or digital form), or other legal act under union or member-state law, setting out key aspects, like the subject-matter and duration of the	As regards the smart contracts, our policies rely on the DataVaults sharing configuration tool to set all the specific characteristics and features of the data sharing (including, for instance, the selection of the privacy/anonymization level, of the visibility level, of the access policies level, of the pricing and of the licensing). The Individual can share not only their

	processing, the nature and purpose of the processing and the types of personal data and the categories of data subjects?	personal data collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase. In this way, whenever an Individual decides to make data available to Data Seekers through the DataVaults platform, they are prompted to define various details regarding data sharing, enabling the construction of fine-grained policies. The various parameters of a data asset sharing configuration can be modified at a later stage by the Individual, who can also completely revoke access to the shared data asset, by appropriately configuring the bound access policies. For the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period. In addition, Andaman7 provides another level of sharing control through its platform to have better control of what can be shared to the DataVaults platform.
Breach Management		
33	Do you have documented personal data breach policies and/or tools for: a) notifying appropriate supervisory authorities of the breach in a timely manner, and with reasons provided for any delays. b) notifying data subjects of high-risk breaches (as defined by GDPR) no later than 72 hours after discovery of a breach, if it is determined (following documented procedures for performing harm risk analysis) that the personal data breach will result in privacy harm to the associated data subjects; and c) Including all items necessary within the notice as required by GDPR?	The policies were developed for the v0.5 and v1.0 of the platform, so after v0.5 we have supported the notification of the data subjects for the case of breaches. Andaman7 does not have any specific documentation related to data breach policy or such tools but falls under its GDPR compliant policy. Tools are already in place if needed.
Security and Privacy by Design		
34	Have the security and privacy protections been embedded into the full lifecycle of automated decision-making processes involving personal data to safeguard the data subject's rights, freedoms and legitimate interests? Have you implemented appropriate technical security and privacy controls, supported by documented privacy principles (e.g., the ISACA Privacy Principles, and/or IEEE privacy standards, etc.), in order to appropriately mitigate harms to individuals to the extent possible in compliance with GDPR and to protect the rights of data subjects?	Yes, this is the main aim of the DataVaults project. Data subjects are able to on-board the platform and control which data they would like to take on-board, then decide if they want to share them, with which data seeker categories, under which provisions, and remove them if they want to. (Unless they have been engaged in a binding contract) Core to these processes are the following components: Data Sharing Configurator; Anonymiser; Access Policy Editor; Data Sharing and Confirmation Dashboard; Risk Management Dashboard; Blockchain Infrastructure; Smart Sharing Contracts. Andaman7 also already provide privacy protections by design in its lifecycle and fully respect GDPR.
Legal, Ethical and Societal Implications		
35	Do people understand your purpose – especially people whose data are about or who are impacted by their use?	The user that wants to share their data to the DataVaults platform is first informed of all the consequences through an information sheet and an informed consent to sign before being able to

	How have you been communicating your purpose? Has this communication been clear? How are you ensuring more vulnerable individuals or groups understand?	configure any exchange of data. We try to present it in the most user-friendly way on the smartphone to ensure that it will be correctly read and understood by the majority. During the demonstration activities, we will ensure that a responsible person can help them, if needed, during this process.
36	What existing ethical codes apply to your sector or project? Are there any legislation, policies, or other regulation shape how you use data, in addition to those listed in DataVaults D2.3? Are you in compliance with them (if applicable)?	Processing and storing of health data falls, of course, under some ethical codes. All are defined in the GDPR, and special country regulation already listed in D2.3. We are in compliance with them.
37	Positive effects on people. Which individuals, groups, demographics or organisations will be positively affected by your DataVaults pilot? How? How are you measuring and communicating positive impact? How could you increase it?	Our main objectives are improving the app to help users to take care of their health and ease collection of data for companies that build a better healthcare system for the future. We then aim to have a positive effect on our users (patients), healthcare companies and ultimately the whole healthcare system. The KPIs should help have a sense of those positive impacts. The analysis of those as well as feedback received during the demonstration will help to improve any positive aspect. We are also communicating regularly and widely on our various social media about the positive impacts of our work.
38	Negative effects on people. Who could be negatively affected by your DataVaults pilot? How? Could the way that data is collected, used or shared cause any harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access to some service? How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.	Negative impacts are not expected by the kind of data collection that is planned during our demonstration. Of course, the fear of sharing such sensitive data will still be present. During the whole demonstration, we are planning to communicate with users in the most transparent way to convince them their data are secured. Any third party involved in the data sharing is involved in the health system and therefore no prejudice or harm should be caused to the data owner. Any report done by a user at any step of the process will be taken into consideration to reveal and fix any harm.
39	Openness and transparency in the design process How open can you be about this project/DataVaults pilot? Are you asking the volunteer for feedback on the project and its outcomes? Are you building in thoughts, ideas and considerations of people affected by your project/DataVaults pilot? How? What information or training might be needed to help people understand data issues?	Any detail about the project will be communicated in the most transparent way at any time of the project. Feedbacks will be collected on a voluntary basis but also according to a predefined time frame in order to get as much feedback as we can get to help us improve our communication and our services. Any feedback will be handled individually and in a transparent way. Help from the technical team behind the DataVaults project may be needed to answer some feedback. In general, we communicate a lot and very transparently on what we do - via our website, our newsletter and our social networks. Please check them.
40	Review and iterations Are you planning to measure, monitor and discuss the data ethics issues in the post-project phase and to review this EDPIA?	Our process regarding ethics in the post-project phase should be the same as described in this EDPIA. Regular reviews will be planned in order to make sure we are still in accord with it.

2.3.1.4 *Demonstrator #4 – Smart home Personal Energy Data*

Introductory questions		
		Responses
1	Nature, scope, context, and purposes of the processing (Recital 90)	The data we are going to collect is mainly consumption data from the homes of the participants. Also, we will collect through questionnaires information about the building they live in and the preferences of the participants in the project. We will only collect this data in order to develop the DataVaults project and the three scenarios of our pilot demonstration. After the end of the Project, only the data related to the Grant Agreement will be kept as required and will only be accessible to the European Commission for a period of 5 years. After that period, the data will be deleted. The purpose of collecting this data is to research new ways to offer services to our clients, using the DataVaults platform as an intermediary to share data between companies and potential clients, in a way that the client will always have control of the data they are sharing. With this data we will be able to recommend energy-saving tips, design PV installations and generate energy demand prediction models.
2	List of in-scope personal data items	1. Consumption data. 2. Data about their home: city, year of construction, m2, kind of roof, orientation, number of people living there. 3. Data about personal preferences: hobbies, communication preferences, tastes, daily habits.
3	List of key data sources	1. MIWenergia's API. 2. Questionnaires through DataVaults platform.
4	Limitation in data source: are there limitations that could influence your project's outcomes? Consider: ● bias in data collection, inclusion/exclusion, analysis, algorithms ● gaps or omissions in data ● provenance and data quality ● other issues affecting decisions, such as team composition	There can be gaps in the consumption data because of meter failures, but these gaps should not affect the project outcomes. Algorithms and data analysis should be carried out carefully to correctly process all the data and obtaining reliable results.
5	Recipients of personal data items. Are you going to be sharing data with other organisations in DataVaults and/or beyond? If so, who? Under what conditions?	All data will be shared through the DataVaults app under the control and consent of the participant. The participant will decide with which company they want to share the data through the app. The data shared with MIWenergia's will only be used to develop our pilot demonstration and will not be shared with any other organisation.
6	Period for which the personal data will be stored (e.g., in hours, days, weeks or years, etc.)	We will only store personal data for the project purposes so when it is finalized, we will delete it.
7	Functional description of the processing operation related to your pilot activities.	When the user wants to connect with MIWenergia source, they need to log in with the MIWenergia credentials. Then they must select the USPC from which they want to share the data and the start and

		end dates. Then, the user's data are being anonymised and encrypted (if the user selects to do so) based on the DataVaults Anonymiser and Encryption tool, based on the anonymisation depth levels selected by the user. User Data can be part of anonymised aggregated "personas" if the user selects his data to be included in those Personas. There his data are merged with data from other people and provided as average data values of a "fictional" individual.
8	What are your main use cases (please provide a summary and indicate the DataVaults deliverable where details can be retrieved)? Are you replacing another product? You can also simply list them (with one line description) and insert the reference to the DataVaults deliverable where the complete description can be retrieved	Our three main use cases are: - Scenario A: PV installation design: Design of a PV installation based on the consumption and the details of their roof. - Scenario B: Improve profiling of the clients to enhance energy efficiency. Profiling of the clients to send them energy tips and improve their energy efficiency. - Scenario C: Energy consumption patterns with personal preferences. Demand prediction and analysis of preferences to provide personalized offers and services. The deliverables in which the details for each scenario are described are "D1.4 Updated DataVaults Concept Definition" and "D6.2 Pilot Scenarios and Implementation Plan". We now use an external application to predict our client's energy demand, which could be substituted in the future with our own model trained with information from DataVaults, if it works better than the actual system.
9	DataVaults technological assets (services, tools…) involved in the data collection and/or processing in your pilot activities	• MIWenergia's API • Data Fetcher • Edge Analytics Engine • Anonymiser • SEAS Analytics Engine
Choice and Consent		
10	How do you provide choices, when appropriate to data subjects regarding use of their personal data? Is consent obtained before using personal information for specific purposes, as required by GDPR? How?	• MIWenergia's participants must sign a consent form before registering in the platform. • Initial Consent is asked upon registration to the platform. However, NO data are shared with anyone at that stage. • Data Subjects have the option to delete their data and/or their account at any time by visiting their profile/account page and asking the system to automatically delete all their data (does not apply to data already acquired through a contract by a data seeker). • The Personal App of DataVaults allows users to fully manage their personal data by deleting it and revoking access to it. (If not already acquired by a data seeker). • Upon sharing data, Users are provided by the Data Sharing configurator with a dialogue requesting them to verify their sharing options of those data.

11	Are consents (once obtained) appropriately documented and maintained? How?	The consent form signed before the start of their participation in the project is stored in MIWenergia's servers. The agreement is stored in the backend database of the Personal App.
12	What data subject's privacy harms might occur? What risk-mitigation measures have you in place?	In DataVaults the user is in total control regarding the sharing of owned data, while the Risk Management dashboard informs data subjects on their privacy exposure. In addition, prior sharing in the cloud platform of DataVaults, data are anonymised and encrypted; this is one of the benefits of DataVaults towards privacy.
Legitimate Purpose Specification and Use Limitation		
13	Are you going to collect only the personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which the data are processed, in support of data-minimization requirements? Is the personal data processing lawful and necessary given the purposes for which the data were collected? Please provide details.	Yes, we will only collect data that is strictly necessary for the purposes of the project. The personal data collection is lawful and necessary to develop the DataVaults project. We will only collect consumption data and other data about their home-building, hobbies and personal preferences through questionnaires in order to develop the three use cases of the project. In every case, the participant will decide which data they want to share with us, giving us permission and will be always informed of what data is shared and with which purpose. The participant will be always in control of the data they share.
Personal Information and Sensitive Information Life Cycle		
14	Are you going to keep personal data for no longer than necessary to support the purposes for which they were collected?	No, we will collect personal data for the project purposes, so only during its deployment.
15	Are you going to collect and/or process one or more of the following types of personal data? If yes, please provide details. a) data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership b) genetic data c) biometric data for the purpose of uniquely identifying a natural person d) data concerning health e) data concerning a natural person's sex life or sexual orientation	No, we are not going to collect any of these types of data
16	How do you ensure that, by default, only personal data which are necessary for each specific processing purpose are processed?	No personal data will be share without consent. The user always chooses which data to make available over the platform. The data subject decides which questionnaires want to answer and share and what not.
Accuracy and Quality		
17	How do you ensure that personal data are kept accurate and up to date, as necessary, and correct personal data errors without delay?	The participant can update their personal data whenever they want through the platform. The personal data is updated following the command of the data subject.
Openness, Transparency and Notice		

18	Are you going to process personal data fairly, transparently and in compliance with applicable legal requirements?	Yes.
19	How do you communicate to data subjects their rights, notices, and answer their questions and provide information to them relating to data processing, in a manner that is clear, easy to understand, and age appropriate to the data subject? Are you going to provide the data subjects, at the time personal data are obtained from them, all necessary information elements? How do you restrict use of their associated personal data?	By default, NO personal data is shared, so this question does not apply. However, once data is shared, we cannot control this, as this is something that falls out of the scope of the project. Maybe we could have some smart contracts which could say something "generic" about "no harm being caused, etc." but again this is not controllable by the system. Regarding MIWenergia demonstrator, its users will have a look and accept the terms of the company. They will also sign the consent form before participating within the project. All relevant information will be provided at the sign-up /registration page, in the relevant terms. An FAQ page will be delivered to answer most common questions and a functional mailbox will be provided for further questions. There will also be a person at MIWenergia available to answer any questions related to the project that the participants want to clarify.
Individual Participation		
20	How are the data subjects allowed to withdraw consent to use their associated personal data at any time, as long as the withdrawal does not result in legal violations about which you have informed the data subjects?	By directly removing their data and/or their account
21	How do you ensure that a data subject can exercise his or her rights mandated by the GDPR? And do you have processes to provide documented reasons for denying requests?	Data Subjects can at any time delete their account and remove their personal data. No requests will be denied. If any participant of the MIWenergia demonstrator asks to leave the project and delete their data, the request will be accepted, and the data removed from our servers. If they have shared their data to other companies through DataVaults, we cannot ensure which data policy the other companies have, but we will help the participant with any procedure.
22	How do you allow the data subject to obtain confirmation regarding whether or not personal data are being processed (including personal data used in partnership with other controllers), and when that is the case, to provide the data subject access to the associated personal data and related mandatory information (including, for instance, information concerning the purposes, the data categories, the recipients, the retention periods, the rights for deletion and registering complaints, etc.) in compliance with GDPR requirements?	Such information is stored in the blockchain where the data shared can be viewed by the data subject. This is provided as part of a dashboard for each user, where they can see: • The data he has already collected but he holds without sharing them. • And the data he has already shared and under which terms (e.g., anonymised or not, for how long, etc.) • The data that (after sharing them) have been acquired by a data seeker and the relevant transaction id.
23	Are you able to provide to the data subject, upon request and without prejudice, a copy of his/her personal data? How? Can this copy be delivered in a commonly used digital format	Yes. Data Subjects can download a copy of the data they have provided to the platform for free. If users need help when downloading, MIWenergia will support them.

	and for free or for a reasonable fee (where the fee is based on actual administrative costs)?	
Accountability		
24	Have you established an appropriately qualified data protection officer (if applicable, according to the GDPR)? Can the data subjects contact him/her for any issue related to processing of their personal data or to the exercise of their rights under GDPR?	The data subjects can contact the project manager at MIWenergia (ana.garcia@miwenergia.com) for any issue related to the processing of their data and their rights and she will try to solve it or ask a capable member of the team to do it. Our DPO can be contacted directly via dpomiwenergia@legitec.com
25	Which DataVaults tool in your demonstrator support accountability?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access
Security Safeguards		
26	How do you ensure that appropriate technical and/or organisational safeguards are implemented to secure personal data, including protections against unauthorized or unlawful processing and against accidental loss, destruction or damage? What DataVaults technologies/tools ensure a level of security for the personal data appropriate to the personal harm risk, including, as appropriate: a) pseudonymisation and/or encryption. b) procedures to establish confidentiality, integrity, availability and resilience of processing systems and services, data backup and recovery; and c) Regular testing of associated security controls?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access • Anonymisation Engine • Data Encryption Engine • TPM Attestation for devices running DataVaults
27	How do you assess the privacy risks and/or the likelihood of privacy harms to data subjects, for instance in the event of unauthorised access, sharing or use of personal data and/or unauthorised or accidental destruction, loss, or alteration of personal data?	Risk Management Dashboard can assess the privacy risks for the shared datasets of the user, based on the sharing policies applied while data are protected by unauthorised access within the platform. Also, with the Blockchain infrastructure for storing transactions, we ensure that no unnoticed alternation, deletion or access will be performed.
Monitoring, Measuring and Reporting		
28	Do you provide reports for data subjects (at specified times; upon their request as appropriate; and reflecting all components required by GDPR), including: a) personal data-breach notifications. b) reports regarding correction of personal data erasure and/or incorrect personal data. c) reports showing the content of personal data associated to the subject. d) reports showing the personal data associated to the data subject shared with others, including the reasons for such sharing; and e) Full digital copies of personal data transmitted directly to another data controller in support of data portability requirements.	a) These are provided upon request via their Inbox page. b) These are provided upon request via their Inbox page. c) Dashboard of collected data. d) Transactions Dashboard.
29	How do you maintain a record of processing activities involving personal data that includes	There is a record showing which data seeker has acquired access to which data set, which is retrieved

	a) the name and contact details for your enterprise/institution; b) the purposes for processing personal data; c) a description of the categories of data subjects and of the categories of personal data involved in the processing; d) the categories of recipients to whom the personal data have been or will be disclosed; and where possible, the established time limits for erasure of the different categories of data?	via the blockchain system and displayed to the user via his sharing dashboard page. Also, information about the prospective categories of data seekers that can have access on each shared dataset is provided in the sharing configuration of each shared dataset, in the access policies lists defined by the data subject.
Preventing Harm		
30	Do you allow the data subjects to request removal of their personal data from automated processing and profiling in situations that could result in adverse legal effects or harms to them?	Yes, in case their data is not already acquired by a third party (as then any data processing happens off the platform and cannot be controlled). See the first two questions under "Individual Participation"
31	How do you ensure that data subjects who exercise their rights (for changing how their personal data are used; request copies of personal data; and/or exercise other rights under GDPR) do not adversely affect the rights and freedoms of others or do not infringe a contract or a binding provision?	The exercise of rights concerns only data not already acquired by another party. If the data is sold to third parties though DataVaults, there will be a contract between the parties engaged in the transaction.
Third-party/Vendor Management		
32	Do you have documented third-party / vendor management policies (and supporting procedures) to specify the type of documented contract (in hard copy and/or digital form), or other legal act under union or member-state law, setting out key aspects, like the subject-matter and duration of the processing, the nature and purpose of the processing and the types of personal data and the categories of data subjects?	As regards the smart contracts, our policies rely on the DataVaults sharing configuration tool to set all the specific characteristics and features of the data sharing (including, for instance, the selection of the privacy/anonymization level, of the visibility level, of the access policies level, of the pricing and of the licensing). The Individual can share not only her/his personal data collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase. In this way, whenever an Individual decides to make data available to Data Seekers through the DataVaults platform, they are prompted to define various details regarding data sharing, enabling the construction of fine-grained policies. The various parameters of a data asset sharing configuration can be modified at a later stage by the Individual, who can also completely revoke access to the shared data asset, by appropriately configuring the bound access policies. For the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period.
Breach Management		
33	Do you have documented personal data breach policies and/or tools for: a) notifying appropriate supervisory authorities of the breach in a timely manner, and with reasons provided for any delays. b) notifying data subjects of high-risk breaches (as defined by GDPR) no later than 72 hours	The policies were developed for the v0.5 and v1.0 of the platform, so after v0.5 we have supported the notification of the data subjects for the case of breaches.

	after discovery of a breach, if it is determined (following documented procedures for performing harm risk analysis) that the personal data breach will result in privacy harm to the associated data subjects; and c) Including all items necessary within the notice as required by GDPR?	
Security and Privacy by Design		
34	Have the security and privacy protections been embedded into the full lifecycle of automated decision-making processes involving personal data to safeguard the data subject's rights, freedoms and legitimate interests? Have you implemented appropriate technical security and privacy controls, supported by documented privacy principles (e.g., the ISACA Privacy Principles, and/or IEEE privacy standards, etc.), in order to appropriately mitigate harms to individuals to the extent possible in compliance with GDPR and to protect the rights of data subjects?	Yes, this is the main aim of the DataVaults project. Data subjects are able to on-board the platform and control which data they would like to take on-board, then decide if they want to share them, with which data seeker categories, under which provisions, and remove them if they want to. (Unless they have been engaged in a binding contract) Core to these processes are the following components: Data Sharing Configurator; Anonymiser; Access Policy Editor; Data Sharing and Confirmation Dashboard; Risk Management Dashboard; Blockchain Infrastructure; Smart Sharing Contracts.
Legal, Ethical and Societal Implications		
35	Do people understand your purpose – especially people who's the data are about or who are impacted by their use? How have you been communicating your purpose? Has this communication been clear? How are you ensuring more vulnerable individuals or groups understand?	Yes, we have explained the purpose of collecting the data for the project and we are always at the disposal of the participants to clarify and explain any issue that they want to ask us. We have been communicating via email and phone calls. This communication has been clear, and they can clear up any concerns at any time via these channels. In addition, all the participants must sign the Informed consent form before participating, so they are aware of the purpose of the project. Vulnerable groups can ask for our support at any time and we will pay special attention to these people to ensure that they always understand the purpose of the data collection and the project.
36	What existing ethical codes apply to your sector or project? Are there any legislation, policies, or other regulation shape how you use data, in addition to those listed in DataVaults D2.3? Are you in compliance with them (if applicable)?	There is no new regulation apart from the one listed in D2.3 that affects the use of the data in the MIWenergia demonstrator.
37	Positive effects on people. Which individuals, groups, demographics or organisations will be positively affected by your DataVaults pilot? How? How are you measuring and communicating positive impact? How could you increase it?	From our energy point of view, people who are responsible for their homes and who are not very educated in energy efficiency and renewable sources will benefit greatly. Also, people interested in reducing their energy consumption and their environmental impact can benefit from the DataVaults deployment. At the same time, if participants reduce their energy consumption or increase their consumption from renewable sources, it will have a positive impact on the environment, reducing CO2 emissions.
38	Negative effects on people. Who could be negatively affected by your DataVaults pilot? How?	We do not consider the users will face negative effects from our collection and use of their data. They will be always aware of what kind of data they are sharing

	Could the way that data is collected, used or shared, cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access to some service? How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.	with us, and we will be supporting them to ensure transparency. The users will face some difficulties just as any other user of an innovative application such as DataVaults. We intend to help them in understanding the use of the platform and make the adaptation process as short as we can. If any user reports any harm, we will take actions to fix and reduce any negative consequences. As the data is anonymised, we do not consider that there is any risk to be re-identified. The data cannot be used to target, profile or restrict access to any people.
39	Openness and transparency in the design process How open can you be about this project/DataVaults pilot? Are you asking the volunteer for feedback on the project and its outcomes? Are you building in thoughts, ideas and considerations of people affected by your project/DataVaults pilot? How? What information or training might be needed to help people understand data issues?	We try to have transparency and openness in the process of the MIWenergia demonstrator deployment. We asked participants for their feedback about the project and platform through questionnaires, in order to improve our developments and take into account their ideas and considerations. We communicate with them in the most transparent way. We have organised workshops with participants to teach them how to use the DataVaults platform and help them to understand the way the data is collected and processed and solve their questions, doubts or issues with the project.
40	Review and iterations Are you planning to measure, monitor and discuss the data ethics issues in the post-project phase and to review this EDPIA?	In the post-project phase, prior to any exploitation activity DataVaults platform by our company, we will review this EDPIA to adapt it to possible changes in legislation or other issues related to data collection.

2.3.1.5 *Demonstrator #5 Personal data for municipal services and the tourism industry*

Introductory questions		
		Responses
1	Nature, scope, context and purposes of the processing (Recital 90)	The processing activities will be carried out inside pilot scenarios to test the functionality the project app and platform that users will adopt to share their personal data with data-seekers, in a secure way and obtaining in exchange a compensation.
2	List of in-scope personal data items	• personal profile data (civil status, mobility preferences, cultural interests, nationality) • social network posts • participation in cultural events • geolocation • Google history • personal certificates
3	List of key data sources	• Profile in DataVaults • Social media • Smartphone • Population registry
4	Limitation in data source: are there limitations that could influence your project's outcomes? Consider:	The main problems faced were in the acquisition process of the required data, before the platform settled, either by accessing the user's smartphone or by connecting external data sources like the

	● bias in data collection, inclusion/exclusion, analysis, algorithms ● gaps or omissions in data ● provenance and data quality ● other issues affecting decisions, such as team composition	population registry, user's position and social networks.
5	Recipients of personal data items. Are you going to be sharing data with other organisations in DataVaults and/or beyond? If so, who? Under what conditions?	Parties acting as data seekers in the pilot activities (Municipality of Prato, Textile Museum, Museum of Palazzo Pretoria, CGIL (Union), CAAF - Fiscal Support Centre) will treat the personal data they will collect according to their own procedures fully respecting national and European legislation.
6	Period for which the personal data will be stored (e.g., in hours, days, weeks or years, etc.)	6 months.
7	Functional description of the processing operation related to your pilot activities.	The user will connect to their own data sources and will be able to share data either non-anonymised or anonymised, according to their own preferences. User's data can also be part of anonymised aggregated "personas", if the user selects their data to be included in those Personas. Here, their data are merged with data from other people and provided as average data values of a "fictional" individual. The data seekers will be able to buy the required assets (either according to the price set buy the data owner, or suggesting a price themselves), once the data owner has accepted the offer.
8	What are your main use cases (please provide a summary and indicate the DataVaults deliverable where details can be retrieved)? Are you replacing another product? You can also simply list them (with one line description) and insert the reference to the DataVaults deliverable where the complete description can be retrieved	• Scenario A - Access to personal data for the analysis of mobility solutions • Scenario B - Access to personal data for the improvement of cultural offer in the city • Scenario C - Access to personal data for the delivery of personal certificates More information is available in D6.2 – Pilot scenarios and Implementation Plan. No replacement of other products is foreseen during the demonstration although some have been identified. And there exists potential to do this.
9	DataVaults technological assets (services, tools…) involved in the data collection and/or processing in your pilot activities.	• Data Fetcher • Edge Analytics Engine • Anonymiser • SEAS Analytics Engine
Choice and Consent		
10	How do you provide choices, when appropriate to data subjects regarding use of their personal data? Is consent obtained before using personal information for specific purposes, as required by GDPR? How?	• Initial Consent is requested upon registration to the platform. However, NO data are shared with anyone at that stage. • Data owners have the option to delete their data and/or their account at any time by visiting their profile/account page and asking the system to automatically delete all their data. (Does not apply to data already acquired through a contract by a data seeker)

		• The DataVaults Personal App allows users to fully manage their personal data by deleting it and revoking access to it. (If not already acquired by a data seeker) • Upon sharing data, users are provided by the Data Sharing configurator with a dialogue requesting them to verify their sharing options of such data.
11	Are consents (once obtained) appropriately documented and maintained? How?	The agreement is stored in the backend database of the Personal App.
12	What data subject's privacy harms might occur? What risk-mitigation measures have you in place?	In DataVaults the user is in total control regarding the sharing of owned data, while the Risk Management dashboard informs data owners on their privacy exposure. In addition, prior to being shared in the cloud platform of DataVaults, data are anonymised and encrypted; this is one of the benefits of DataVaults towards privacy.
Legitimate Purpose Specification and Use Limitation		
13	Are you going to collect only the personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which the data are processed, in support of data-minimization requirements? Is the personal data processing lawful and necessary given the purposes for which the data were collected? Please provide details.	The pilot's activities aim at validating the functionalities of the DataVaults platform and the model proposed for sharing and remunerating personal data: only strictly necessary data will be acquired and processed, in accordance with GDPR and other regulation requirements.
Personal Information and Sensitive Information Life Cycle		
14	Are you going to keep personal data for no longer than necessary to support the purposes for which they were collected?	We store it only for the scope of the project.
15	Are you going to collect and/or process one or more of the following types of personal data? If yes, please provide details. a) data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership b) genetic data c) biometric data for the purpose of uniquely identifying a natural person d) data concerning health e) data concerning a natural person's sex life or sexual orientation	No
16	How do you ensure that, by default, only personal data which are necessary for each specific processing purpose are actually processed?	The data owners are the only one able to decide which data will be made available over the platform. Data Seekers will be only able to buy data made available by data owners.
Accuracy and Quality		
17	How do you ensure that personal data are kept accurate and up to date, as necessary, and correct personal data errors without delay?	Personal data are fully in a data owner's control, and they can decide the updating schedule according to needs.
Openness, Transparency and Notice		
18	Are you going to process personal data fairly, transparently and in compliance with applicable legal requirements?	Yes
19	How do you communicate to data subjects their rights, notices, and answer their	By default, no personal data is shared, so this question does not apply. However, once data is

	questions and provide information to them relating to data processing, in a manner that is clear, easy to understand, and age appropriate to the data subject? Are you going to provide the data subjects, at the time personal data are obtained from them, all necessary information elements? How do you restrict use of their associated personal data?	shared, we cannot control this, as this is something that falls out of the scope of the project. Maybe we could have some smart contracts which could say something "generic" about no harm being caused, etc. but again this is not controllable by the system. Regarding the demonstrators, users will be informed on their rights and data processing through a specific information form available on the platform and they will be able to provide their consent. Each demonstrator will ensure the compliance with GDPR and other national regulation, according to its own data management procedures. All relevant information will be provided at the sign-up /registration page, in the relevant terms. An FAQ page will be delivered to answer most common questions and a functional mailbox will be provided for further questions.
Individual Participation		
20	How are the data subjects allowed to withdraw consent to use their associated personal data at any time, as long as the withdrawal does not result in legal violations about which you have informed the data subjects?	By directly removing their data and/or their account.
21	How do you ensure that a data subject can exercise his or her rights mandated by the GDPR? And do you have processes to provide documented reasons for denying requests?	Data Subjects can at any time delete their account and remove their personal data. The Municipality of Prato is familiar with this kind of personal data processing operations: personal data processing is already used by the administration in different contexts. It has already in place technical and organizational measures for risk assessment and risk control in relation to data protection: register of treatments and risk assessment procedures when necessary, according to the GDPR (art.35), appointment of DPO and definition of appropriate policies for the correct behaviour of employees. No requests will be denied.
22	How do you allow the data subject to obtain confirmation regarding whether or not personal data are being processed (including personal data used in partnership with other controllers), and when that is the case, to provide the data subject access to the associated personal data and related mandatory information (including, for instance, information concerning the purposes, the data categories, the recipients, the retention periods, the rights for deletion and registering complaints, etc.) in compliance with GDPR requirements?	Such information is stored in the blockchain where the data shared can be viewed by the data subject. This is provided as part of a dashboard for each user, where he can see: • the data he has already collected but he holds without sharing them. • And the data he has already shared and under which terms (e.g., anonymised or not, for how long, etc.) • The data that (after sharing them) have been acquired by a data seeker and the relevant transaction id.
23	Are you able to provide to the data subject, upon request and without prejudice, a copy of his/her personal data? How? Can this copy be delivered in a commonly used digital format and for free or for a reasonable fee (where the fee is based on actual administrative costs)?	Yes, the data owner can download a copy of the data they have provided to the platform for free.

Accountability		
24	Have you established an appropriately qualified data protection officer (if applicable, according to the GDPR)? Can the data subjects contact him/her for any issue related to processing of their personal data or to the exercise of their rights under GDPR?	Yes, the Municipality of Prato has in place technical and organizational measures for risk assessment and risk control in relation to data protection, including the appointment of DPO who may be contacted on request.
25	Which DataVaults tool in your demonstrator support accountability?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access
Security Safeguards		
26	How do you ensure that appropriate technical and/or organisational safeguards are implemented to secure personal data, including protections against unauthorized or unlawful processing and against accidental loss, destruction or damage? What DataVaults technologies/tools ensure a level of security for the personal data appropriate to the personal harm risk, including, as appropriate: a) pseudonymisation and/or encryption. b) procedures to establish confidentiality, integrity, availability and resilience of processing systems and services, data backup and recovery; and c) Regular testing of associated security controls?	• Blockchain infrastructure for storing transactions. • Access Policy Engine for safekeeping access • Anonymisation Engine • Data Encryption Engine • TPM Attestation for devices running DataVaults
27	How do you assess the privacy risks and/or the likelihood of privacy harms to data subjects, for instance in the event of unauthorised access, sharing or use of personal data and/or unauthorised or accidental destruction, loss, or alteration of personal data?	Risk Management Dashboard can assess the privacy risks for the shared datasets of the user, based on the sharing policies applied while data are protected by unauthorised access within the platform. Moreover, through the Blockchain infrastructure for storing transactions, we ensure that no unnoticed alteration, deletion or access will be performed.
Monitoring, Measuring and Reporting		
28	Do you provide reports for data subjects (at specified times; upon their request as appropriate; and reflecting all components required by GDPR), including: a) personal data-breach notifications. b) reports regarding correction of personal data erasure and/or incorrect personal data. c) reports showing the content of personal data associated to the subject. d) reports showing the personal data associated to the data subject shared with others, including the reasons for such sharing; and e) Full digital copies of personal data transmitted directly to another data controller in support of data portability requirements.	a) These are provided upon requests via their Inbox page. b) These are provided upon requests via their Inbox page. c) Dashboard of collected data d) Transactions Dashboard e) Transactions Dashboard
29	How do you maintain a record of processing activities involving personal data that includes a) the name and contact details for your enterprise/institution; b) the purposes for processing personal data; c) a description of	There is a record showing which data seeker has acquired access to which data set, which is retrieved via the blockchain system and displayed to the user via their sharing dashboard page.

	the categories of data subjects and of the categories of personal data involved in the processing; d) the categories of recipients to whom the personal data have been or will be disclosed; and where possible, the established time limits for erasure of the different categories of data?	Moreover, information about the prospective categories of data seekers that can have access on each shared dataset is provided in the sharing configuration of each shared dataset, in the access policies lists defined by the data owner.
Preventing Harm		
30	Do you allow the data subjects to request removal of their personal data from automated processing and profiling in situations that could result in adverse legal effects or harms to them?	Yes, in case their data is not already acquired by a third party (as then any data processing happens off the platform and cannot be controlled). See the first two questions under "Individual Participation"
31	How do you ensure that data subjects who exercise their rights (for changing how their personal data are used; request copies of personal data; and/or exercise other rights under GDPR) do not adversely affect the rights and freedoms of others or do not infringe a contract or a binding provision?	The exercise of rights concerns only data not already acquired by another party.
Third-party/Vendor Management		
32	Do you have documented third-party / vendor management policies (and supporting procedures) to specify the type of documented contract (in hard copy and/or digital form), or other legal act under union or member-state law, setting out key aspects, like the subject-matter and duration of the processing, the nature and purpose of the processing and the types of personal data and the categories of data subjects?	As regards the smart contracts, our policies rely on the DataVaults sharing configuration tool to set all the specific characteristics and features of the data sharing (including, for instance, the selection of the privacy/anonymisation level, of the visibility level, of the access policies level, of the pricing and of the licensing). The Individual can share not only their personal data collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase. In this way, whenever an Individual decides to make data available to Data Seekers through the DataVaults platforms, they are prompted to define various details regarding data sharing, enabling the construction of fine-grained policies. The various parameters of a data asset sharing configuration can be modified at a later stage by the Individual, who can also completely revoke access to the shared data asset, by appropriately configuring the bound access policies. For the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period.
Breach Management		
33	Do you have documented personal data breach policies and/or tools for: a) notifying appropriate supervisory authorities of the breach in a timely manner, and with reasons provided for any delays. b) notifying data subjects on high-risk breaches (as defined by GDPR) no later than 72 hours after discovery of a breach, if it is determined (following documented procedures for performing harm risk analysis) that the personal data breach will result in	The policies were developed for the v0.5 and v1.0 of the platform, so since v0.5 we have supported the notification of the data subjects for the case of breaches.

	privacy harm to the associated data subjects; and c) Including all items necessary within the notice as required by GDPR?	
Security and Privacy by Design		
34	Have the security and privacy protections been embedded into the full lifecycle of automated decision-making processes involving personal data to safeguard the data subject's rights, freedoms and legitimate interests? Have you implemented appropriate technical security and privacy controls, supported by documented privacy principles (e.g., the ISACA Privacy Principles, and/or IEEE privacy standards, etc.), in order to appropriately mitigate harms to individuals to the extent possible in compliance with GDPR and to protect the rights of data subjects?	Yes, this is the main aim of the DataVaults project. Data subjects are able to on-board the platform and control which data they would like to take on-board, then decide if they want to share them, with which data seeker categories, under which provisions, and remove them if they want to (unless they have been engaged in a binding contract) Core to these processes are the following components: Data Sharing Configurator; Anonymiser; Access Policy Editor; Data Sharing and Confirmation Dashboard; Risk Management Dashboard; Blockchain Infrastructure; Smart Sharing Contracts
Legal, Ethical and Societal Implications		
35	Do people understand your purpose – especially people whose data are about or who are impacted by their use? How have you been communicating your purpose? Has this communication been clear? How are you ensuring more vulnerable individuals or groups understand?	Yes, we have explained the purpose of collecting the data for the project and we are always at the disposal of the participants to clarify and explain any issue that they want to ask us. We have been communicating via email and phone calls. This communication has been clear, and they can clear up any concerns at any time via these channels. In addition, all the participants must sign the Informed consent form before participating, so they are aware of the purpose of the project. Whilst we have not engaged with vulnerable groups, they can ask for our support at any time and we will pay special attention to these people to ensure that they always understand the purpose of the data collection and the project.
36	What existing ethical codes apply to your sector or project? Are there any legislation, policies, or other regulation shape how you use data, in addition to those listed in DataVaults D2.3? Are you in compliance with them (if applicable)?	The use of data in the demonstrator will be fully compliant with current legislation policies and regulations.
37	Positive effects on people. Which individuals, groups, demographics or organisations will be positively affected by your DataVaults pilot? How? How are you measuring and communicating positive impact? How could you increase it?	Positive effects might arise both for all groups of citizens and data seekers (municipality, cultural institutions, and fiscal service providers) in the improvements of public and cultural services and in the facilitation in the management and sharing of personal certificates. Moreover, we assume that the demonstrator will contribute in arising citizens' awareness on personal data management risks and will empower users in maintaining the control on their data sharing. We have access to survey tools to measure such impacts and to disseminate results through social channels and communication media active in the city.
38	Negative effects on people.	We are not foreseeing any negative effects or harm on people and the technical solutions put in place in the

	Who could be negatively affected by your DataVaults pilot? How? Could the way that data is collected, used or shared cause any harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access to some service? How are limitations and risks communicated to people? Consider: people whom the data is about, people impacted by its use and organisations using the data.	DataVaults tools will reasonably prevent the possibility of individuals' re-identifications. Nevertheless, had the survey tools which were adopted in the open demonstration phase had highlighted negative impacts perceived by the users. specific countermeasures would have been put in place accordingly.
39	Openness and transparency in the design process How open can you be about this project/DataVaults pilot? Are you asking the volunteer for feedback on the project and its outcomes? Are you building in thoughts, ideas and considerations of people affected by your project/DataVaults pilot? How? What information or training might be needed to help people understand data issues?	In the pilot activities users' opinion was taken into account by activating some feedback process related with the demonstrator activities, for example through specific surveys but also social media campaigns managed by the administration. The topic of personal data sharing is increasingly important, and some specific communication and training actions should be carried out through social media and institutional channels. At the moment no specific communication tools are included in the DataVaults app/platform, but should this happen, we will consider this extra opportunity.
40	Review and iterations Are you planning to measure, monitor and discuss the data ethics issues in the post-project phase and to review this EDPIA?	The data ethics issue is a very relevant topic for the Municipality of Prato, and it will be included in the administrative debate whenever required, also in accordance with current national and European legislation. No specific review of this EDPIA is foreseen after the conclusion of the project, although the document might provide useful insights for the debate on the data ethics issue.

2.3.2 Collection of Citizens' feedback

The DataVaults Personal App User Interface was perceived in a positive manner regarding its capabilities to allow the user to exercise real control over their personal information. However, in a scale from 1 (minimum) to 5 (maximum), a relevant number of respondents selected 3 or 4. Several respondents felt comfortable when using DataVaults as regards the exercise of their rights (information, access, rectification of inaccurate data, erasure, object, etc.), though a small number was not satisfied in this regard. The DataVaults Personal App was considered enough easy-to-use for adjusting one's privacy preferences. The evaluation of the clarity of the DataVaults notification regarding one's privacy risk exposure was good: several respondents selected 4 in a scale from 1 (minimum) to 5 (maximum). On the other hand, the consent mechanism in DataVaults is perceived still somehow fatiguing, since in many replies the average value in a scale from 1 (minimum) to 5 (maximum) was 3. In general, the project's technology was considered as privacy and data protection friendly, with most of the respondents selecting 5 in a scale from 1 (minimum) to 5 (maximum). However, in some case (Prato) the average value in a scale from 1 (minimum) to 5 (maximum) was 3.

2.3.3 Fulfilment of the Ethical, Legal, Privacy, Security and Trust Requirements

As already underlined in D6.3, the legal and ethical requirements for the design, development and validation of DataVaults cloud-based platform and Personal App, as well as, to some extent, for the future operation of them, are elicited and refined/updated respectively in D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management" and D2.3 "Updated DataVaults Security Methods and Market Design". They were elicited, relying on the Fairness & Privacy-by-Design-and-by-Default Approach enriched with the Protection Goals method, as well as on the analysis of the regulatory landscape and on the factual analysis of the privacy-relevant properties and other features foreseen for the personal data collection, processing and sharing in DataVaults services and tool.

The same deliverables also set the technical security, privacy and trust requirements for the DataVaults platform and App, elaborated with a view on an enhanced (holistic) data sharing solution. This category of requirements was elicited reflecting on DataVaults work and data flow and on how data security, privacy, sharing and management services are to be engrained in a policy-compliant Blockchain structure.

Both two macro-categories of requirements include mandatory and desirable/recommended requirements.

The fulfilment of the requirements regarding the DataVaults technology ensures that it is legally compliant, ethically sound and gives rise to a trusted, secure privacy-friendly enhanced (holistic) data sharing solution.

The following Table 3, which does not cover other "usability" aspects which will be covered as non-functional specifications/requirements, shows the progress made between the interim assessment and the final evaluation of the fulfilment of the Legal, Ethical, Privacy, Security and Trust requirements set in D2.1 and refined in D2.3. The result at the Final evaluation being that all of them were met by the DataVaults technological artefacts.

Requirement	Status	Alpha	Beta	Final
Legal requirements to be fulfilled				
Purpose limitation and legitimate aim	Yes **Completed**	DataVaults Personal App Terms and Privacy Notes DataVaults Cloud Platform Terms and Privacy Notes		
Data minimisation	Yes **Completed**	DataVaults Personal App Terms and Privacy Notes DataVaults Cloud Platform Terms and Privacy Notes Access Policy Engine	Deletion of unnecessary information coming from data sources Revised data sharing templates	Risk Privacy Metrics Dashboard and Engine
Data Accuracy	Yes **Completed**		TPM DAA Module Component	
Integrity and Confidentiality	Yes **Completed**	Anonymise Access Policy Engine	Persona Builder	ABE Engine
Storage Limitation	Yes **Completed**			DataVaults Secure Storage Containers

Transparency	Yes **Completed**	Access Policy Engine	DataVaults DLT infrastructure Sharing Configuration Information Checklist	
Privacy and Data Protection by Design and Privacy by Default	Yes **Completed**		Persona Builder	DataVaults Secure Storage Containers ABE Engine
Avoidance of discrimination (including social sorting) and of harm	Yes **Completed**	DataVaults Personal App Terms and Privacy Notes DataVaults Cloud Platform Terms and Privacy Notes		
Informed Consent	Yes **Completed**	DataVaults Personal App and Cloud Platform Terms and Privacy Notes		
Set of requirements referring to the voluntary participation to DataVaults demonstrators	Yes **Completed**	DataVaults Personal App Terms and Privacy Notes DataVaults Cloud Platform Terms and Privacy Notes	DataVaults Mobile App Terms and Privacy Notes	
User Control	Yes **Completed**	Access Policy Engine		Risk Privacy Metrics Dashboard and Engine Personal Wallet Sharing Wizards
Data subject's rights	Yes **Completed**	Access Policy Engine Data Request Resolver	DataVaults DLT infrastructure	Risk Privacy Metrics Dashboard and Engine Personal Wallet
Enforcement	Yes **Completed**		DataVaults DLT infrastructure Access Policy Engine	
Fairness by Design	Yes **Completed**	The whole design and development are driven by this approach.		
Effective "sharing the wealth" paradigm	Yes **Completed**	The whole design and development are driven by this approach.		
Privacy Notice	Yes **Completed**	DataVaults Personal App Terms and Privacy Notes DataVaults Cloud Platform Terms and Privacy Notes		
Data breaches	Yes **Completed**			DataVaults Personal App Penetration Test DataVaults Cloud Platform Penetration Test
Accountability	Yes **Completed**		Trusted DLT Engine	DataVaults Operations Manual
Record of processing activities	Yes **Completed**			DataVaults Operations Manual (Videos)

Data Protection Impact Assessment	Yes **Completed**			Risk Privacy Dashboard and Engine
Technical and organizational measures	Yes **Completed**	Access Policy Engine		DataVaults Secure Storage Containers ABE Engine
Use of private environment/cloud as much as possible	Yes **Completed**			DataVaults Secure Storage Containers ABE Engine
User and data protection friendly User Interface	Yes **Completed**			Risk Privacy Dashboard and Engine Sharing Wizards
Measures in case of profiling	Yes **Completed**		Persona Builder	Risk Privacy Dashboard and Engine Personal Wallet
Appointment of Data Protection Officer	Yes **Completed**	DPO Assignment		
Assignment of responsibilities	Yes **Completed**	Responsibilities Assignment		
Ethics Board set-up and involvement	Yes **Completed**	Ethics Board Setup		
Checklist for the Privacy, Security and Trust				
Integrity and Confidentiality	Yes **Completed**	Sharing Configurator Component		
Authorization and Access Control	Yes **Completed**	Access Policy Engine		
Non-repudiation and Accountability of Actions	Yes **Completed**		Trusted DLT Engine	
Anonymity	Yes **Completed**	Not for the alpha stage version	Persona Builder	
Conditional Anonymity	Yes **Completed**	Anonymiser		
Unlikability	Yes **Completed**		Trusted DLT Engine	
Data Privacy	Yes **Completed**	Anonymiser	Persona Builder	ABE Engine
Forward and Backward Privacy	Yes **Completed**			Trusted DLT Engine
Trustworthiness and Operational Correctness	Yes **Completed**	Personal Wallet		
Cryptography	Yes **Completed**	Sharing Configurator		ABE Engine
Ledger Security	Yes **Completed**	Trusted DLT Engine		

Table 3: Checklist for Legal, Ethical, Privacy, Security and Trust requirements

2.4 CONCLUDING REMARKS

The Layer I and the Layer II of the Legal, Ethical, Security, Privacy and Trust Evaluation demonstrate that both the ethics-and-privacy-by-design-and-by-default approach, as described in D9.2, and the requirements elicited in D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management" and enriched in D2.3 "Updated DataVaults Security Methods and Market Design", have been properly followed by the technical team of the project during the design and development of the DataVaults technologies and by the demonstrators during the validation phase. The DataVaults Ethics Advisory Board in its last meeting confirmed that the Consortium has worked so that the first version of the platform and its demonstrators reach the necessary conditions for legal, ethical and user friendliness. It is clear that progress can still be made to reach an even better balance between legal / ethical constraints and user friendliness / user comprehension. This will be a gradual process that can be expected from future production systems.

The DataVaults demonstrators' processes and operations adequately implemented the ethical procedures outlined in D9.2 and in WP10 deliverables. Furthermore, the demonstrators strictly adhered to the Ethical and Legal requirements specifically set for them in the mentioned deliverables, such as the need to follow adequate consent procedures and recruiting procedures and to use suitable tools and safeguards.

The Ethics and Data Protection Impact Assessments, conducted by the demonstrators, were particularly useful for evaluating the legal, ethical, security, privacy and trust aspects of the DataVaults technology (platform and app), including the assessment of their components, tools and services in the demonstrators, therefore in different application contexts.

The remarks and findings of such EDPIAs are also expected to be relevant for the post-project phase, when the projects' outcomes will be taken up and used in real-life environments. In this regard, also the lessons learnt and blueprints for DataVaults' uptake and operation (as reported in D8.6) are expected to play a significant role.

These considerations, and in particular the outcomes of the EDPIAs and the fulfilment of the requirements regarding the DataVaults technology elicited in D2.2 and D2.3, confirm that the DataVaults platform, its app and the other technological assets generated in this research are legally compliant and ethically sound and give rise to a trusted, secure privacy-friendly and citizen-respectful data sharing environment.

3 FINAL STATUS OF THE DEMONSTRATION SITES

This chapter presents the final status of the demonstrators operating in the DataVaults projects, providing the complete view from the beginning of their operation to the end of the project, alongside with the scenarios realised.

For this cause, in each demonstrator the different scenarios chosen (and which have been presented in deliverable D6.2) are once again briefly described to give to the reader the complete picture, and thereafter details about the realisation and the evolution of each scenario are provided.

It is noted that the technical evaluation of the demonstrators regarding the DataVaults platform is covered in chapter 5 of this document.

3.1 DEMONSTRATOR #1 - SPORTS AND ACTIVITY PERSONAL DATA (OLYMPIACOS)

Olympiacos Sports Club has a large base of members and fans who register or renew their subscription on an annual basis. The satisfaction of the members & fans are very important issues for the club as they are the most important source of revenues. Moreover, as the club has different sport departments at the competitive level and the academies, it needs to manage a large base of professional and young athletes, including important contact, activity and medical details.

The DataVaults demonstrator focused on the utilisation of the platform to allow the club to give the opportunity to these target groups to collect and share their data directly with the club, to allow the latter to start building added value services for them.

3.1.1 Target Audience Reached during Final Phase

3.1.1.1 *Data Owners*

The main data owners that were relevant to the Olympiacos demonstrator were people from the fanbase of the club, as well as athletes of the club. During the execution of the different scenarios, stakeholders of both those groups were reached.

Regarding individuals categorised as “fans”, those included people from the club’s fanbase, some club employees (on boarded in the previous stages of the demonstrator for performing internal testing activities and for getting trained to use DataVaults). At the end of M40 of the project, 493 fans have been registered in the DataVaults platform.

Regarding the other target audience to act as data owners, that of athletes, during the demonstrator phase 55 athletes have been registered in the DataVaults platform.

As such, the overall count of data owners onboarded in the Olympiacos case has been of 548 individuals, with approx. 70% of them being fans and 30% athletes.

3.1.1.2 *Data Seekers*

During the previous phases of the demonstration and in the final one, the sole seeker was Olympiacos, who engaged with the data owners identifies above for retrieving data from some of them and deliver the relevant compensation.

Contacts with different club sponsors have been made, as well as with the professional club departments (e.g. Football club, Basketball club as well as the Volleyball club), which showed an interest to join the platform once an "Olympiacos" deployment of the platform is made available (after the end of the project), while initial discussions for some other DataVaults connectors (for the professional club departments) have taken place, to identify other fan data that those entities hold ((such as donations to those clubs, past ticketing/attendance activity, info relevant of banking products linked to fans (Olympiacos FC Visa card, etc.) and that could be made available to enrich the profiles of individuals.

3.1.2 OLYMPIACOS Demonstrator Scenarios Evolution

This section provides a status update on the validity and progress of the initial scenarios defined for the demonstrator in D6.2

It is noted that all objectives of the scenarios listed below have been reached, except from the branding of the PersonalApp, as effort was put to localise the platform in Greek. Minor branding activities have been performed, mostly in terms of "colouring palette" selection for the demonstrator, preparing the ground for the future branding of the platform, which is covered in D7.4.

3.1.2.1 *Scenario A - Club Fans and Members Personal Data Marketplace*

3.1.2.1.1 Scenario Description

In this scenario, fans and members of the club are encouraged to collect their personal data, understand what it comprises, manage and share them through the DataVaults platform. For the fans and members, awareness on how the type and quantity of personal data relates to compensation offered from the club shall encourage the former to push forward sharing and keeping up these data. For the club, this will help build a stronger relationship/interaction with the fans and members, understand their needs, and offer better services, as well as incentives to share more of their personal data, respond to the requests that may come from different organizations and find new sponsors/partnerships.

The main goal of this scenario is to collect information related to social media activity and the preferences/likes of individuals who are already fans and members of the club. A secondary goal would be to collect location data related to the position of the individual, for example if they are present at the stadium during the time the club's team is playing. Collecting and analysing such data could assist the club to combine personal data already existing in its systems, with social media and location activity to engage and interact better with the fans and members.

In principle, the main objectives of this scenario remained the same. However, as there was a limitation in collecting social media data (due to different social media like Facebook or Twitter prohibiting data collection via an API for using this data within the project), it was

decided to bring forward the secondary objective of the scenario which dealt with location data and combine it with the profile information of data owners. As such, a new objective has been inserted in the table below (see final row).

3.1.2.1.2 Objectives of OLYMPIACOS Scenario A

Objective	Status	Phase Achieved
Connect internal CRM of the club to DataVaults cloud platform.	Achieved	Final
Share data of fans and members with DataVaults cloud platform.	Achieved	Final
Brand the DataVaults PersonalApp, for example as an affiliate of Olympiacos.	Postponed	See D7.4
Promote DataVaults PersonalApp to existing fans and members (e.g., via e-mail).	Achieved	Final
Advertise and implement compensation for connecting the DataVaults app with social media data sources based on the features that will be offered by the platform.	Achieved	Final
Push questionnaires to fans to extract information in the absence of other data sources.	Achieved	Final
Fetch data from Mobile App sources as well as other data	Achieved	Final

Table 4: OLYMPIACOS Scenario A objectives

3.1.2.1.3 OLYMPIACOS Scenario A Evolution

The initial steps for this scenario had to do with the identification of the data that was held within the CRM of the club to check the possibility to port these data into DataVaults as a dataset owned by a fan, allowing him to retrieve his (and only his) data, which were currently only stored by the club. Once the data categories were analysed, certain ones which were considered valuable for the scenario have been selected, which resolved around the attendance of fans in different events, the “spending power” they potentially have, as well as other aspects such as how much money they spend in the club’s shops (like boutiques, etc). These data, have been considered as important for the club and for relevant sponsors, and the idea for retrieving them and making them available by the fans themselves would allow interested stakeholders (as well as the club itself) to utilise them in various analyses and try to improve their service offerings (as the already collected data were not allowed to be used directly by those entities due to the terms under which they have been collected).

An overview, shown as a relationship diagram of such data categories that have been selected for retrieval from the CRM are shown in the next figure.

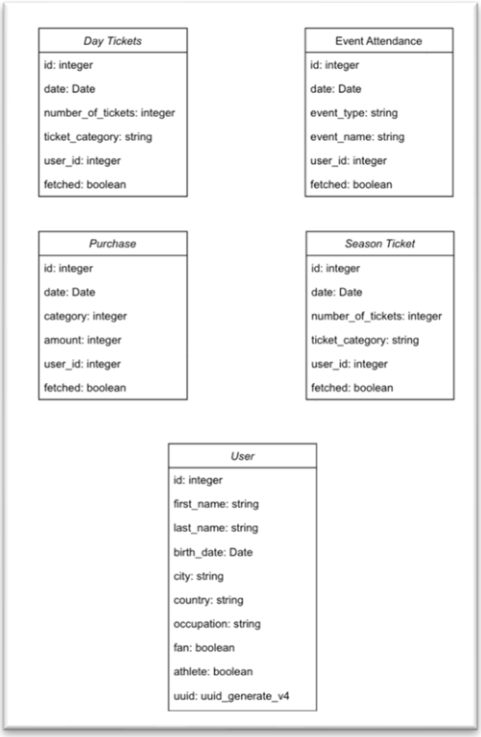

Figure 1: Example of Data Extracted for each Individual from the CRM

Following this analysis, the realisation of the connector has been developed, allowing individuals to connect to the CRM by using their credentials (combination of their memberID number and their birth data) to retrieve their own data and develop such an asset.

Moreover, two rounds of questionnaires have been sent to fans of the club, asking them their opinion regarding some activities of the club relevant to fans experience during and after different events, accumulating in this manner valuable information regarding the preferences of fans around aspects that do also consider sponsors and collaborators of the club, to be able to present them with this information to improve the fans experience. The first questionnaire has been answered by 88 individuals in total, as this has been pushed to fans at M34 of the project, while another version of the questionnaire has been pushed to individuals at M38 of the project and 176 responses were collected.

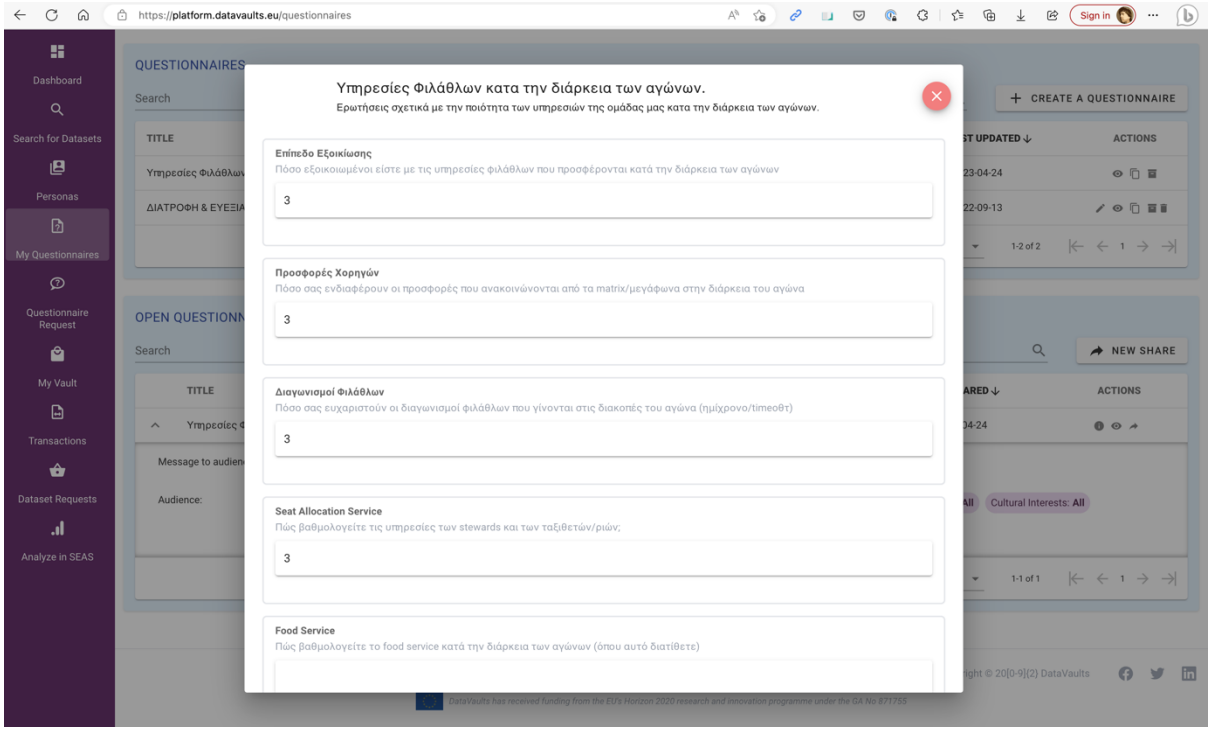

Figure 2: Example of Questionnaire “Experience During Events” sent to Individuals.

Localisation of the overall platform has been also performed in collaboration with the technical partners, in order to allow individuals to work in an environment supporting the Greek language, which was considered essential, while initial work on the “branding” activities has taken place, working on the “colour palette” that seemed more relevant to the club.

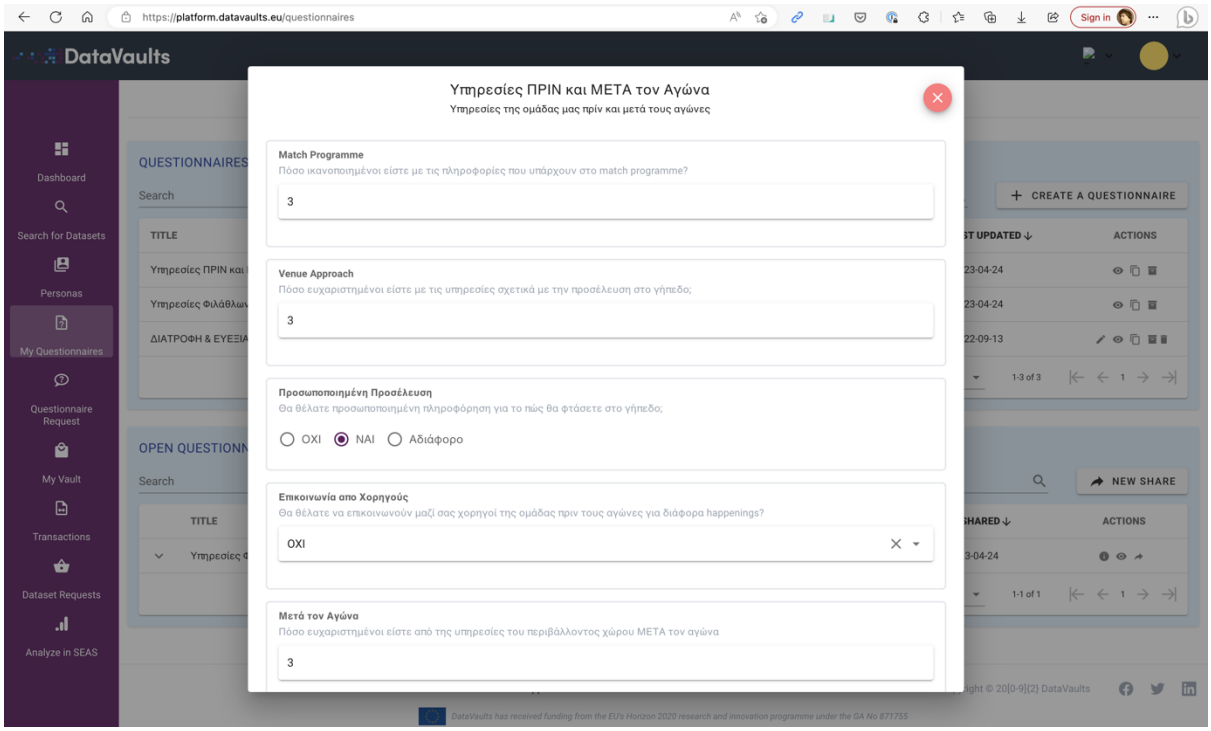

Figure 3: Example of Questionnaire “Experience Before and After Events” sent to Individuals.

At the same time, as an extra data source relevant for the demonstrator, the DataVaults Mobile App has been utilised. The Mobile App proved valuable to identify location relevant information (especially the route tracking) for combining these with the other information of individuals to design some reports that would be interesting to the club and to sponsors.

For this purpose, the existence of the Mobile App has been communicated by the club to individuals during selected events, and at the end of the demonstrator's phase 98 individuals have download the App, while 84 have shared their data (either routes or activity data), through the app at least one time during the overall demonstration period. Figure 10 presents examples of simple routes performed by an individual after a specific event held at the Karaiskakis football field or Peace and Friendship Stadium, both located in the area of Piraeus/Faliro, and which are the home ground of the Olympiacos Football, Basketball and Volleyball teams.

The overall scenario was concluded by the club acquiring data from individuals (utilising the features of the DataVaults platform) and providing the relevant compensation to those individuals.

Data sources

As identified above, the main data sources used for Scenario A were the following:

- Data Coming from the club's owned CRM, where the most crucial information about each individual has been made available to that individual (data owners) through a dedicated connector operating within DataVaults.
- Mobile App data source, focusing on route tracking that presented the whereabouts of data owners prior and/or after events.
- Questionnaires sent to the data owners via the DataVaults Questionnaire facility.

Scenario Findings

The data acquired during the execution of this scenario allowed the extraction of valuable information for the synthesis of the club's fan base and their preferences, as well as some statistics about them.

Regarding statistics, from the sample of the 218 people who have shared their CRM data, various conclusions drawn from them are presented in the following figures. It is considered that the pool of data owners included in this analysis is representative for the complete fan base, nevertheless the more people onboarded, the more fine-grained the results would be.

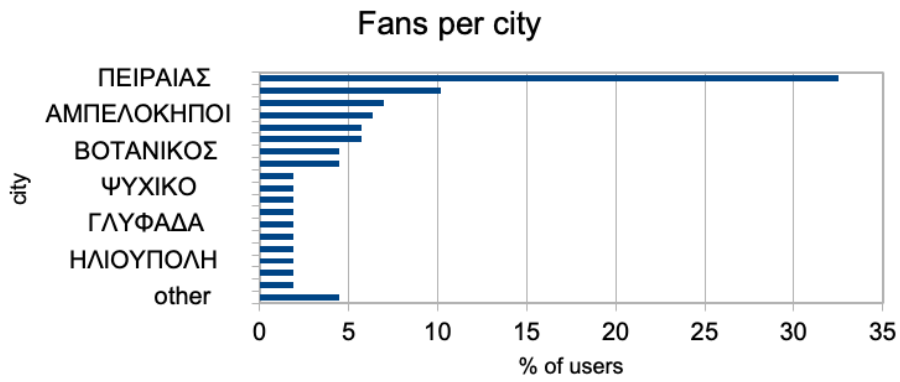

Figure 4: Fans Per City Registered as Fans (Percentage)

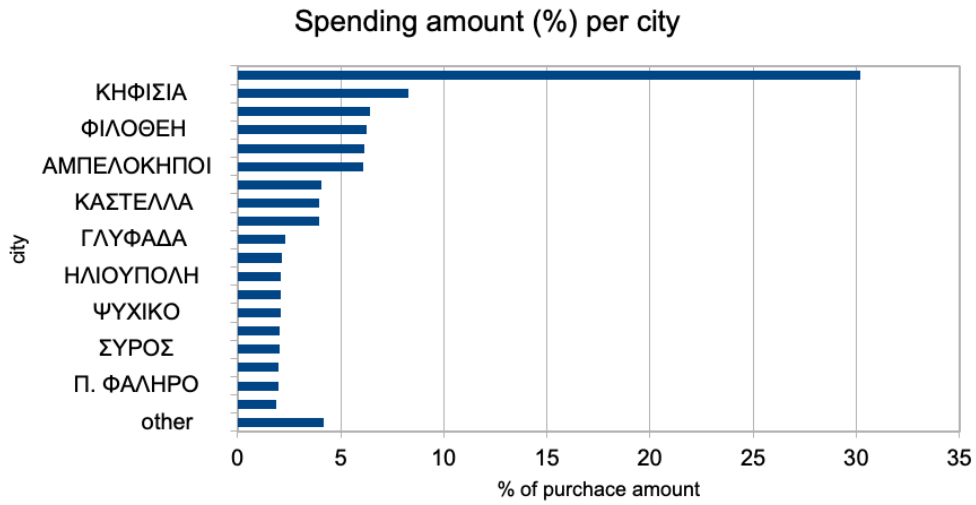

Figure 5: Spending Amount (%) per City

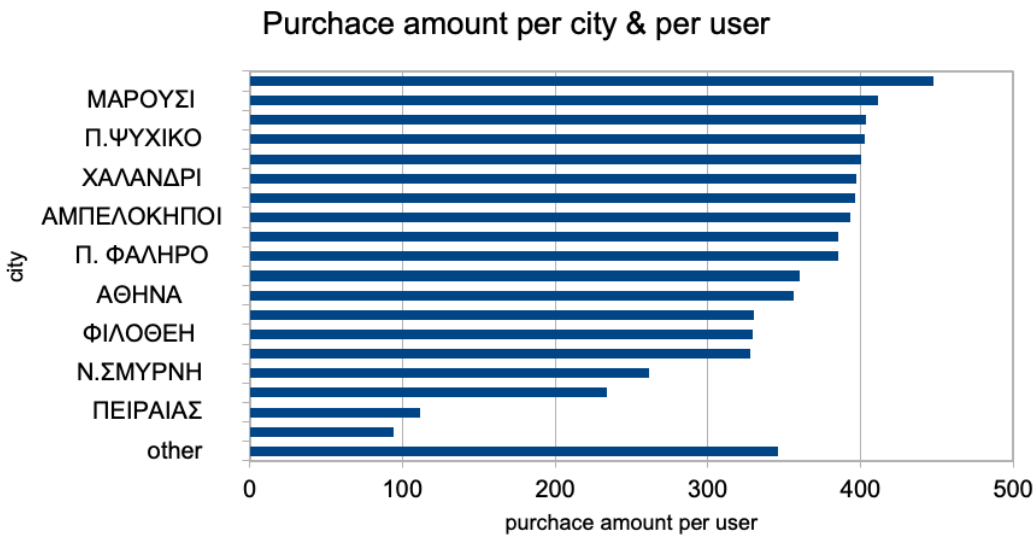

Figure 6: Relating Spending Amounts per User to the Cities they live in

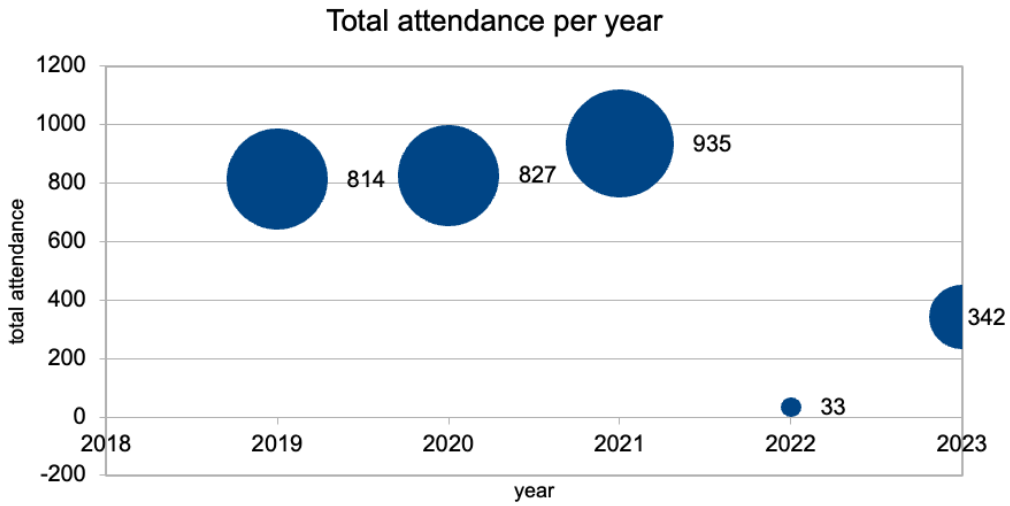

Figure 7: Total Attendance per Year

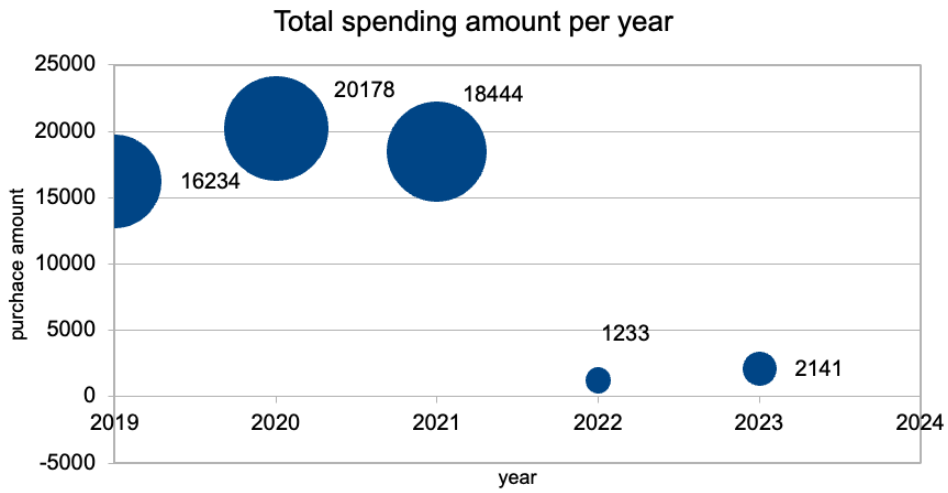

Figure 8: Total Spending Amounts (of the users providing their data) per Year.

Apart from the data that do provide a more detailed and combined view of data owners by combining the CRM data, also an analysis of the different questionnaires has been performed to understand the degree of satisfaction of fans relevant to the services offered before, during and after the different sport events attended by the fans.

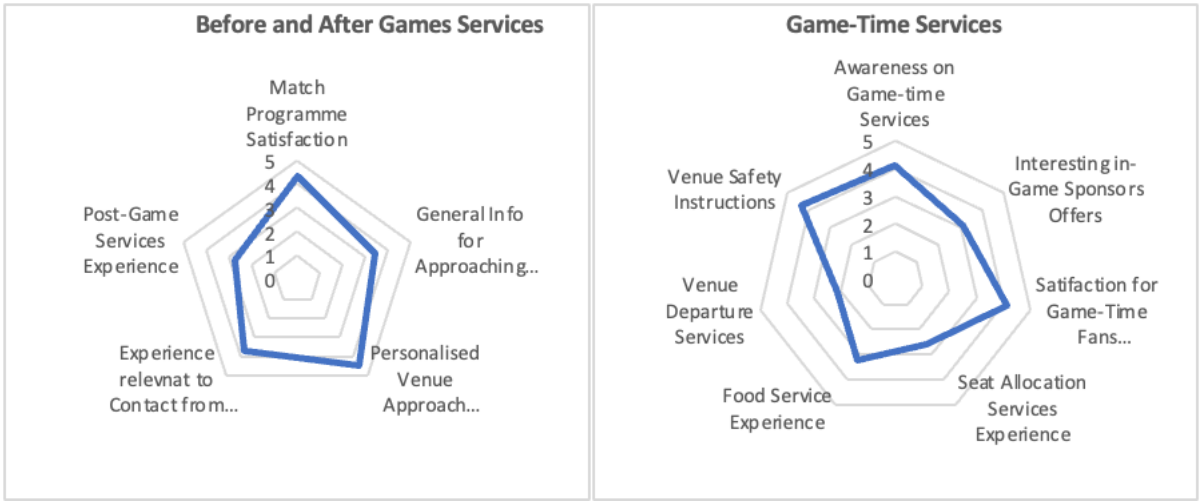

Figure 9: Responses to Questionnaire for acquiring feedback relevant to games experience.

This exercise proved very useful for acquiring direct feedback from fans, as this is at the moment is performed manually (by Olympiacos employees asking attending fans and filling in forms), and is something that is quite hard to do due to the fact that in most of the cases fans are reluctant to provide this information, possibly due to personal data considerations but also as time is limited and fans are anxious to reach the venue (or leave in a hurry).

Moreover, the analysis of the different routes recorded via the Mobile App has also provided very valuable information relevant to the whereabouts of fans mostly following events (as most of the routes were tracked following events). As the routes that were shared over DataVaults were from less users than the overall pool of data owners (e.g. 69 data owners shared routes, with approx. 3 entries each), it was decided to classify routes into four different categories, three based on the type of season ticket held by the users (e.g. high-end, middle tier, low-end season ticket prices), and an additional one for holders of single game tickets. In this context, the “fans” profiles were linked to the routes recorded by themselves, and an analysis took place to identify on a map the whereabouts of those individuals.

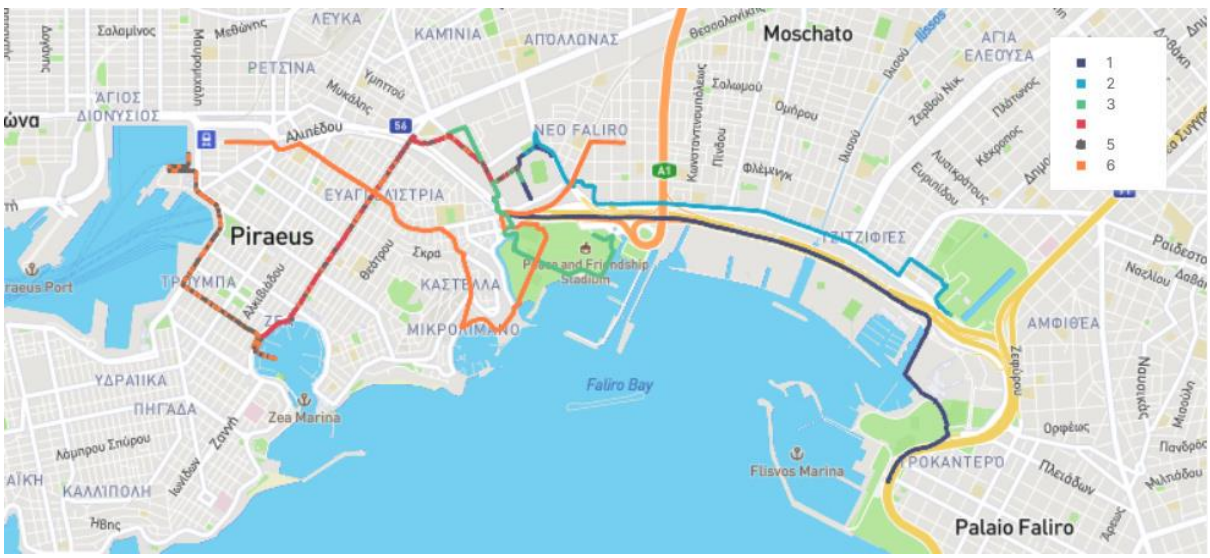

Figure 10: Example of Single Routes tracked by Individuals (of different groups) visualised over the SEAS environment using Superset.

By retrieving all those datasets, they were sorted into three different categories and merged to see the main waypoints of people belonging to the same group. As the graphs below suggest, holders of single event tickets tend to the nearby region of Mikorlimano or the Stavros Niarchos Park, which are both within walking distance. The first area is an area with many restaurants and coffee shops, while the second area is an open-air park situated by the sea. Based on these data, it becomes apparent that many people that choose to visit an Olympiacos event as a single occasion (e.g., not frequent audience/season ticket holders), tend also to spend some hours in nearby shops and do not immediately leave the area surrounding the venue.

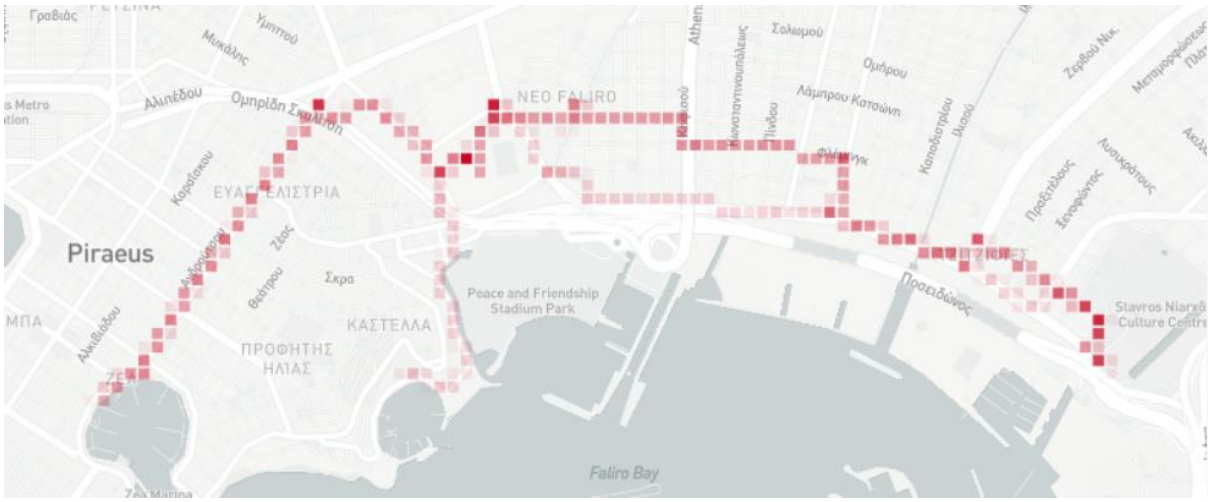

Figure 11: Heatmap of Whereabouts Single Ticket Holders

In contrast, most owners of season-tickets of low prices tend to walk directly to the metro station or the bus stops, signalling that most of this audience directly leaves the venue and does not spend much time on the surrounding area.

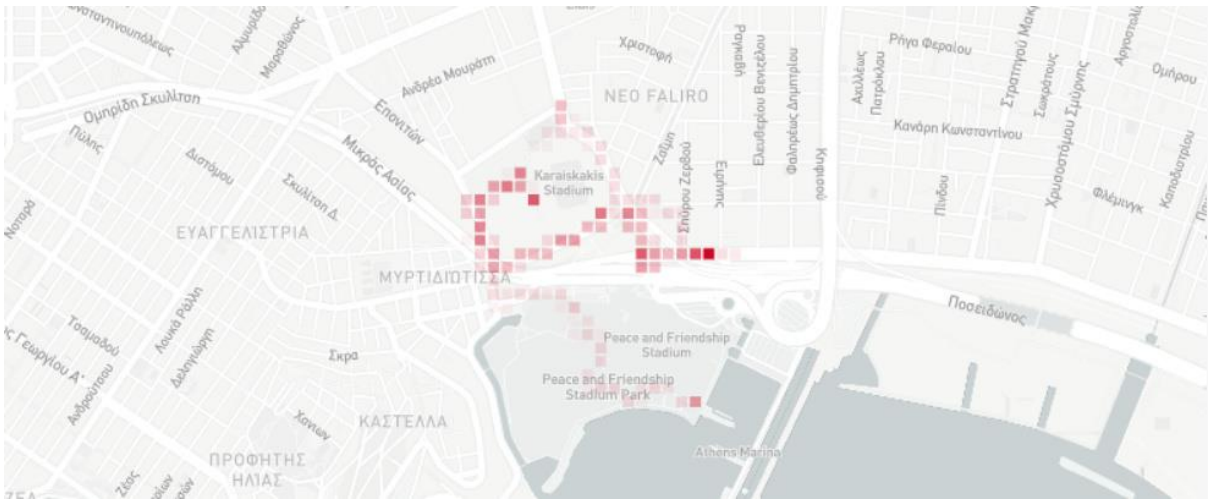

Figure 12: Heatmap of Whereabouts of Low-Tier Price Seasonal Tickets

At the same time, when visualising routes from people holding middle-tier price seasonal tickets, it is evident that many of them choose to walk along the coastline and visit the Floisvos

marina (more than the Mikrolimano area), which is an area with more expensive restaurants and shops.

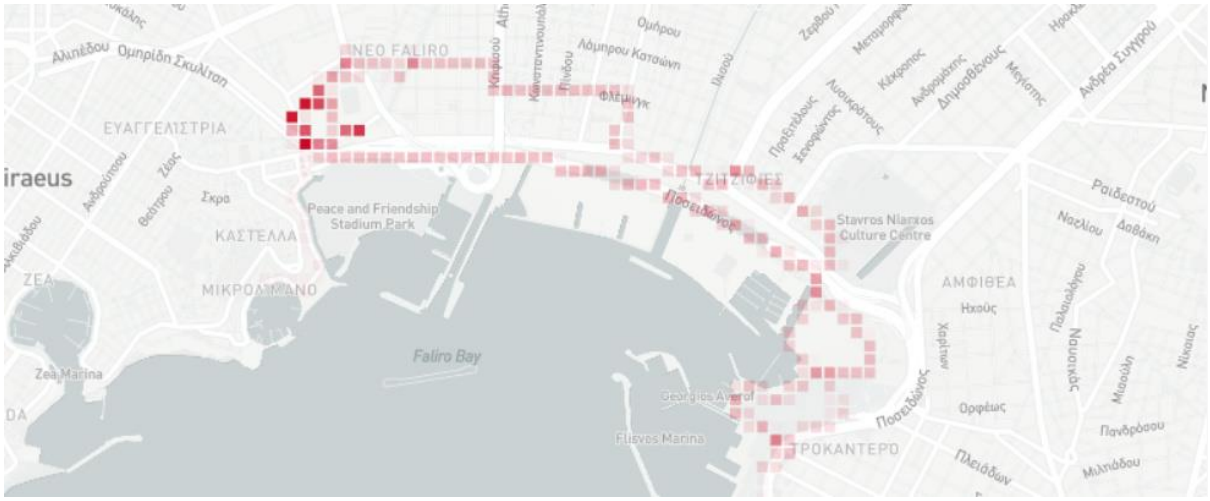

Figure 13: Heatmap of Whereabouts of Middle-Tier Price Seasonal Tickets

Finally, when it concerns high-end price level ticket holders, no routes have been recorded, despite such individuals having installed the app on their mobile phones.

3.1.2.2 *OLYMPIACOS Scenario B – Athletes Sports and Activity Data Sharing*

3.1.2.2.1 Scenario Description

In this scenario, athletes of the club were encouraged to share their personal data, athletic activity data and ergometric and medical examination data through the DataVaults platform.

The main goal of this scenario is to collect especially athletic activity data and ergometric and medical examination data of athletes who already belong to the club. Collecting and analysing such data could assist the club since the current portal on the one hand includes name, surname, address, telephone number, e-mail, date joined the club, sports (e.g., tennis, sailing, etc.), but on the other hand it doesn't allow the entry of more complicated data such as the results of the ergometric and medical examinations.

3.1.2.2.2 Objectives of OLYMPIACOS Scenario B

Despite bringing the scenario close to its realisation, the last objective as shown in the table below was postponed, as it was not possible at the moment to utilise DataVaults as the core channel where all athletes' data could be collected.

Objective	Status	Phase Achieved
Connect internal portal of the club to DataVaults to fetch information upon individuals' command.	Achieved	Final
Share personal data and athletic activity data of athletes with DataVaults cloud platform.	Achieved	Final
Promote DataVaults PersonalApp to athletes, trying to include a proportion of professional athletes.	Achieved	Final
Run analytics on the performance of the athletes based on the collected data.	Achieved	Final

Use DataVaults platform as the main channel to collect athletic activity and ergometric and medical examination data from the athletes who take part in the initiative.	Postponed	See D7.4

Table 5: OLYMPIACOS Scenario B objectives

3.1.2.2.3 OLYMPIACOS Scenario B Evolution

For realising this scenario, work has initially been placed in identifying various data sources and repositories where data from athletes is stored. This investigation, which was performed in some selected sports department in the beginning, resulted in a list of sources and repositories that signalled that there would be an issue (in terms of resources) for collecting all these data.

In this context, the final objective as shown in the table above was not met, mainly for two reasons. First, such data, especially ergometric data, which are considered sensitive require the athlete to give approval for their collection. As part of athletes joining the club, such an activity is performed in the first place, enabling the club to collect these data and store them in the registries belonging to each sports department. However, in case these need to be transferred to DataVaults, new consent procedures must be designed and executed, which would have a direct impact on the current way of working in the club. Secondly, as identified before, most of the athlete's data reside in different repositories, as the club has many sports departments and there is no central database at the moment. Going one step further, it is the case that even for the same individuals data are stored as well in different repositories/formats, as for example training statistics are stored in manual .xls, game statistics are stored directly in the repositories of the statistics companies, while biometrics and medical examination data are stored at the premises of the medical partner of the specific sports department of club, etc.), while many of those data are not accessible online. As a result, in order to be able to allow athletes to collect all their data in one place (e.g., under their DataVaults account), the integration efforts would be required, which were not within the scope of the project.

For this reason, it was agreed that selected athletes would voluntarily join DataVaults, and upload data coming from activity trackers they own (e.g., through Google Fit (using the DataVaults Mobile App) or through Fitbit), as well as by uploading some data stored in files, to generate data assets that would include personal activity metrics.

Data sources

As identified above, the main data sources used for Scenario B were the following:

- Mobile App data source, for personal activity metrics recorded in Google Fit
- Fitbit connectors data, for personal activity metrics
- Direct upload of file including activity/vitals metrics

Scenario Findings

As identified above, the scope of this scenario did not focus on analyses of data from athletes, as these are handled by specialised sports software that are customised to measure specific parameters important for each sport activity.

3.1.3 High Level Impact Recorded in the Demonstrator

3.1.3.1.1 Benefits for the Data Seekers recorded from the Scenario.

The sole data seeker during the demonstration phase was Olympiacos itself.

Through the use of the platform, Olympiacos, as a data seeker, came to a position to reach a number of fans that up until now were reached only via telephone contact or via questionnaires filled in at the surrounding area of the club's stadium prior to games, and could retrieve valuable information about them. This information concerned different dimensions, such as their economic spending capacity/history (coming through the CRM of the club), their attendance in games, their routes around the stadium during match days, etc. Using all this data, Olympiacos (as data seeker) was able to better understand specific patterns of the fanbase (even at the small scale that the experiment was conducted) and become aware of information which has not been known to the club, as it was very hard to get access to such resources.

In the athletes' scenario, the main benefits had to do with collecting some of the online and scattered information at a central place, especially of ergometric and biometric data coming directly from athletes, and that are not today fully available to the club in a systematic and structured online manner. Although this data collection was not possible due to the technical issues identified above, the collection of activity tracking data is also considered beneficial for the club and the first step towards the generation of such a unified "databases" system, where the data is owned by the athletes, and they share it with the club as well.

Judging from the experience gained via the demonstration phase, it became evident that by using such a platform, data seekers will be in a position to re-use information that has been initially collected for another purpose, provided that the data owners retrieve them (from the systems already stored via various methods) and place them back into a personal data marketplace such as DataVaults, while at the same time they would be able to access and combine data coming from diverse sources (as it happens in scenario A).

Moreover, another direct benefit for data seekers is the ability to run some analyses by deploying the analytics infrastructure provided by DataVaults, without the need for them to choose and install another system for working with the data.

3.1.3.1.2 Benefits for the Individuals recorded from the Scenarios.

Benefits for data seekers in the frame of the demonstrator were identified at two different levels.

Direct benefits had to do with the provision of compensation (redeemable to specific items provided by the demonstrator). It was recorded (via the questionnaires) that the provision of such compensation greatly impacted the will of stakeholders to share their data. Moreover, another direct benefit has been the fact that individuals could receive more targeted questionnaires (without having their identity disclosed), to provide feedback to the club.

Indirect benefits to data owners have to do with the provision of improved services to data owners, based on the analysis of their data as performed by the data seeker (in this case the

club). Whilst this benefit was not actually recorded during the demonstration phase, however it is the main driver to achieve the long-term impact KPIs for the club, judged by improving the services and the overall experience to the individuals which share their data.

3.1.3.1.3 Benefits for the Organisation recorded from the Scenario.

During the demonstration phase, it became evident that Olympiacos has recorded a number of benefits from the operation of the DataVaults platform which can be categorised in the following paragraphs.

First, the club was able to introduce a new way to manage the data of its members and fanbase, re-using existing data that has been stored in the CRM of the club, and combining it with new data coming directly from stakeholders. This is a very valuable resource for the operation of the club, as it allows it to push forward for the realisation of the greater vision of providing more services to both fans and other stakeholders, by working on rich data that revolve around the behaviour and the profile of its clients (e.g., the fans/members). This will allow the club to open up new revenue streams, by providing improved services to this audience, as well as providing insights to other entities, such as sponsors, local authorities, etc, as such actors are in great need of knowledge relevant to the preference of fans regarding mobility and commuting options (for example the case of municipal and traffic management authorities, or of local businesses), regarding spending power (for example the case of sponsors and advertising companies), etc.

Second, the access to athletes' data, in a more holistic manner could also greatly benefit the club, as the different departments will be in a position to have access to more data-rich profiles of athletes, and better organise their internal activities, opening the scene for introducing data science in various sports (as done by some professional teams around Europe), to improve the overall training activities and allow for more personalised training services to be offered to athletes.

Third, using the methods provided by DataVaults, the club is in a position to directly access data owned by individuals, without the many limitations that GDPR imposes, as it is the fans that choose to provide their data to the club, so the latter comes to a position where it can use the acquired data in a legitimate manner, respecting data owners privacy, as it is not re-using data collected for other causes (for example the already stored data in the CRM).

3.1.3.1.4 Demonstrator KPIs

In this section, we present the KPIs relevant to the demonstrator as identified in deliverables D6.1 and D6.2, as well as other KPIs recorded during the demonstration phase.

It is noted, that not all these KPIs could be measured during the demonstrator phase of the project, as some of them concern post-project targets, as part of the exploitation activities of the demonstrator. Those are marked as such "Long-Term Impact" KPIs and are discussed in deliverable D7.4

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Status at M40	Comments
Increase in Stakeholder Trust	Initial survey showed 65% level in trust. Number of club's stakeholders who are "satisfied" and "very satisfied" with the club's participation in the DataVaults initiative as a percentage of club's stakeholders who completed the satisfaction survey	Targeted to achieve 80% level in trust.	Survey questions responded to by members and fans who participate in DataVaults initiative.	98%	Target Achieved
More effective management of members and fans data	Current satisfaction level of 50% Number of members & fans who are "satisfied" and "very satisfied" with the club's methods of handling personal data as a percentage of members & fans who completed the satisfaction survey	Targeted to achieve 75% satisfaction level.	In-house mechanisms	94%	Target Achieved
More effective management of sport activity data.	Current level of effectiveness deemed to be 60% Number of athletes who are "satisfied" and "very satisfied" with the club's methods of handling personal data as a percentage of athletes who completed the satisfaction survey	Targeted to achieve 75% level of effectiveness.	In house analytics. Satisfaction surveys completed by athletes who participate in DataVaults initiative	100%	Target Achieved
Fans profiling	N/A at the moment due to GDPR Correlation of fans' profile data with spending and attendance	Complete Profile of Fans	By accessing the "profile" data asset shared by each fan	Complete Profile of Fans	Target Achieved
Whereabouts of fans after events	Heatmaps generated based on fans routes. N/A	Identify 2 main routes	Analysis of routes classified per fan ticket "profile"	3 main routes	Target Achieved
Access to daily activity data of athletes	N/A	1 personal activity dataset per athlete	Collection of Ergometric files provided by athletes	1 Ergometric online profile	Target Achieved

	Correlation of athletes' profile with activity tracking data			dataset per athlete	
More registered members	Current level of 130,000 registrations Number of members who have renewed their membership with the club for the current season	Targeted to achieve 135,000 registrations.	In house database count.	N/A	Long-Term Impact KPI Discussed in D7.4
More "active members"	Currently 20,000 active members on an annual basis- Number of members who have participated in the club's advertised activities.	Targeted to achieve 22,000 active members.	In house data. In-house analytics (e.g., records of members who took part in the annual general assembly of the club)	N/A	Long-Term Impact KPI Discussed in D7.4
More "active fans"	Currently 60,000 (on an annual basis) Number of fans who have followed the club's games in more than one sports department.	Targeted to achieve 70,000 "active fans"	In house analytics.	N/A	Long-Term Impact KPI Discussed in D7.4
Increased sponsorship revenue	Currently stands at €800,000 per annum	Targeted to achieve €1,000,000 per annum.	In house data.	N/A	Long-Term Impact KPI Part of D7.4

Table 6: Demonstrator #1 KPIs

3.1.4 Demonstrator's Activities Timeline

The following table presents the main activities performed within the demonstrator during the period under consideration.

Demonstrator 1 OLYMPIACOS	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40
Scenario A - Club Fans and Members Personal Data Marketplace																						
Alpha Phase																						
Connect internal CRM to DataVaults																						
Share data of fans and members																						
Collect data from a few of early adopters																						
Club stakeholders inspect data																						
Final phase																						
Open up the use of DataVaults personal app																						
Acquire analytics from the DataVaults cloud platform																						
Pilot a branded version of the DataVaults personal app																						
Final Phase																						
Branded DataVaults app available to everyone																						
Activate the sharing compensation mechanisms																						
Acquire analytics from the DataVaults cloud platform																						
Club stakeholders verify personas																						
Invite club sponsors																						
Scenario B - Athletes Sports and Activity Data Sharing																						
Alpha Phase																						
Transform static isolated data																						
Connect data sources to DataVaults																						
Test sharing features																						
Club stakeholders inspect data																						
Final phase																						
Showcase basic analytics extracted with DataVaults																						
Recruit (professional) athletes																						
Final phase																						
Branded DataVaults app available to everyone																						
Acquire analytics from the DataVaults cloud platform																						
Showcase acquisition of athletic activity and ergometric and medical examination data																						
Invite club sponsors																						

Table 7: Execution Timeline for Demonstrator #1 – OLYMPIACOS

3.2 DEMONSTRATOR #2 - STRENGTHENING ENTREPRENEURSHIP AND MOBILITY (PIRAEUS)

Piraeus is the largest & busiest port in Greece, among the biggest European ports and the main hub connecting Europe, Asia and Africa. Piraeus is the maritime and former industrial centre of the Athens metropolitan area, and one of the most densely populated cities in Europe. Apart from commercial trading, the traffic of cruise passengers is very remarkable, causing equally heavy traffic in the road network of the Piraeus centre.

Through DataVaults, PIRAEUS aspires to renovate the way it collects, manages and analyses data, moving from simple information provision and limited feedback solutions towards embracing novel citizens and visitors' engagement practices. The DataVaults app will allow the targeted stakeholders to provide more and richer data to the city, entailing more personalised services, always in a trusted and secure manner, respecting their privacy.

3.2.1 Target Audience Reached at Final phase.

The following sub sections provide a status update on the audience reached by the demonstrators following the final phase of the demonstrator.

3.2.1.1 *Data Owners*

The Municipality of Piraeus has established a list of over 150 data owners, as a group consisting of Municipal personnel and citizens derived from PireApp, which is the Municipal application for declaring problems within the city. From that pool of participants, we had 92 people who were involved in Scenario A, 89 participants in Scenario B and 135 participants in Scenario C.

3.2.1.2 *Data Seekers*

Following the final phase of the demonstrator the Municipality of Piraeus is the sole data seeker. Initial contact (in person meeting) with the Piraeus Trade Association has been established, a stakeholder in one of our scenarios and a potential future data seeker in the DataVaults platform.

3.2.2 PIRAEUS Demonstrator Scenarios Evolution

This section provides a status update on the validity and progress of the initial scenarios defined for the demonstrator in D6.2

It is noted that all objectives of the scenarios listed below have been reached, except from the branding of the PersonalApp, as this was expected to have minimal impact on the demonstrator scenarios and more effort was required to be rerouted by the technical partners to improving the operation of the platform.

3.2.2.1 *PIRAEUS Scenario A - Smart Mobility Services for Individuals*

3.2.2.1.1 Scenario Description

The first scenario aimed to get insights towards better scheduling the mobility strategy and the relevant services within the city. The main goal in this scenario was to collect all relevant

information from citizens and fans, especially routes of motion around the sport venues and information of attendance to specific sport events and get insights relevant to road traffic diversions in order to avoid traffic congestion around the sport venues and the entrance roads towards the city of Piraeus. The collected number of datasets is somehow less than initially expected, due to the limited time in running the final phase of the demonstrator and therefore openly engaging participants. However, the results coming out of the analysis are close to the expected outcomes of the pilot.

3.2.2.1.2 Objectives of PIRAEUS Scenario A

Objective	Status	Phase Achieved
Setup the DataVaults PersonalApp to contain the necessary for this scenario information	Achieved	Final
Extract data of citizens from existing apps (e.g., PireApp) to be used for promotional purposes, always following the agreement/instruction of the user to retrieve their data	Achieved	Beta
Brand the DataVaults PersonalApp, to be able to promote it through the Municipality of Piraeus channels.	Postponed	See D7.4
Advertise and implement compensation (e.g., perks) for connecting the DataVaults app with social media data sources (e.g., Facebook profile, Instagram profile, etc.) based on the features that will be offered by the platform	Achieved	Final
Push questionnaires to citizens and fans in order to extract information in the absence of other data sources	Achieved	Final

Table 8: PIRAEUS Scenario A: objectives

3.2.2.1.3 Scenario A: Evolution during Final phase

The main goal in this scenario is to collect all relevant information from citizens and fans, especially routes of motion around the sport venues (Karaiskaki Football Ground and Peace & Friendship Stadium) at times of specific sport events. The main goal is to establish through data analytics the best road traffic diversions and parking guidelines in order to avoid traffic congestion around the sport venues and the entrance roads towards the city of Piraeus.

The initial step for Scenario A was to collect citizen's information and form a basic group of users, using data from Pireapp, the Municipal app used to collect citizen's complaints, ideas and suggestions, also used as one of the main channels for informing citizens about what happens in Piraeus. Using Pireapp information we have reached out to citizens and tourists, promoted DataVaults and managed to form a final group of over 150 users for our pilots.

The second step was to demonstrate to all participants the use of the DataVaults PersonalApp and of the DataVaults mobile app. Several sessions were conducted in order to demonstrate the use of the PersonalApp and provide accurate information on how to use the mobile app in times of sport events, in order to actually capture real routes of fans going to the sport venues. The participants in these sessions were asked to expect the questionnaires that the Piraeus pilot was planning to deploy to obtain the required personal data.

Questionnaires

The next step for us was to develop these questionnaires, based on the information and the experience we had from the paper questionnaires previously deployed to establish baselines KPI values, but also from the finalised SUMP (Sustainable Urban Mobility Plan) that the Municipality of Piraeus has been establishing in the past year. For this, we sought the assistance of the Department of Programming and Development, the responsible Department for the Municipality of Piraeus SUMP.

The aforementioned questionnaires were created within the DataVaults app, finalised and deployed to the Piraeus pilot participants. It consisted of 7 questions, 2 demographical and 5 topics related. A value of 50 points was given for answering the questionnaire. At the time of downloading the answers to perform our analysis, 92 people had answered the Scenario A questionnaire.

Mobile app

Several of the above users also used the mobile app provided, in order to record routes at time of events. From the mobile app a set of 34 routes were recorded during an important football game on the 23rd of April 2023. These datasets were obtained for 80 points.

Analysis of questionnaires and routes

The amount of data collected through the use of questionnaires within the DataVaults platform, and the smaller amount of information collected through the deployment and use of the DataVaults mobile app were analysed using the SEAS toolbox. A traffic heat map at times of sport events was created based on this information. The results verified our expectations and came in line with the findings and suggestions of the city's SUMP.

3.2.2.2 *PIRAEUS Scenario B - Empowering local entrepreneurship*

3.2.2.2.1 Scenario Description

In this scenario, the data to be provided by the DataVaults users will be used to better understand consumer behaviours and preferences, with the aim to strengthen the local economy through activities that can be brought forward by the municipality. The main goal in this scenario was to collect all relevant information from citizens and tourists, especially commercial interests and preferences. The collected number of datasets was somehow less than initially expected, due to the limited time in running the final phase of the demonstrator and therefore openly engaging participants. The results coming out of the analysis have been fed to the Piraeus Trade Association, as well as to the local Destination Management Organization to be taken into consideration for promotional activities during the upcoming Piraeus Sea Days festival.

3.2.2.2.2 Objectives of PIRAEUS Scenario B

Objective	Status	Phase Achieved
Setup the DataVaults PersonalApp to contain the necessary for this scenario information	Achieved	Final

Extract lists of citizens from existing apps (e.g., PireApp or an upcoming touristic application) to be used for promotional purposes	Achieved	Beta
Promote DataVaults PersonalApp to citizens (e.g., via e-mail or through the website) and tourists (e.g., via posters and leaflets at the Cruise Terminal)	Achieved	Final
Advertise and implement compensation (e.g., perks) for connecting the DataVaults app with data sources based on the features that will be offered by the platform	Achieved	Final
Push questionnaires to citizens and tourists in order to extract information in the absence of other data sources	Achieved	Final

Table 9: PIRAEUS Scenario B objectives

3.2.2.2.3 Scenario B: Evolution during Final phase

The main goal in this scenario is to collect information from citizens and tourists, especially their commercial interests and preferences, in order to establish through data analytics personalised promotional actions to enhance the local commerce and entrepreneurship in the city of Piraeus. Also, the data provided by the DataVaults users will be used to better understand consumer behaviours and preferences, with the aim to strengthen the local economy through activities that can be brought forward by the Municipality. Moreover, Piraeus has contacted the Piraeus Traders Association as an immediate stakeholder to join the platform and act as 2nd tier data seeker.

The initial step for Scenario B was to collect citizen's information and form a basic group of users, using data from Pireapp, the Municipal app used to collect citizen's complaints, ideas and suggestions, also used as one of the main channels for informing citizens about what happens in Piraeus. Using Pireapp information we have reached out to citizens and tourists, promoted DataVaults and managed to form a final group of over 150 users for our pilots.

The second step was to demonstrate to all participants the use of the DataVaults PersonalApp and of the DataVaults mobile app. Several sessions were conducted in order to demonstrate the use of the PersonalApp and provide accurate information on how to use the mobile app in times of sport events, in order to actually capture real routes of fans going to the sport venues. The participants in these sessions were asked to expect the questionnaires that the Piraeus pilot was planning to deploy in order to obtain the required personal data.

Questionnaires

The next step for us was to develop these questionnaires, based on the information and the experience we had from the paper questionnaires previously deployed to establish baselines KPI values, under the guidance of the Piraeus Destination Management Organization.

The aforementioned questionnaires were created within the DataVaults app, finalized and deployed to the Piraeus pilot participants. It consisted of 8 questions, 2 demographical and 6 topics related. A value of 40 points was given for answering the questionnaire. At the time of downloading the answers in order to perform our analysis, 88 people had answered the Scenario B questionnaire.

Analysis

The amount of data collected through the use of questionnaires underwent analysis using the SEAS toolbox and spreadsheets. The results were presented to the Destination Management Organization and the Piraeus Trade Association to be taken into consideration for promotional activities during the upcoming Sea Days festival (more than 100 events taking place in Piraeus between the 27 May and the 5 June).

3.2.2.3 *PIRAEUS Scenario C - Services for Personalized cultural and touristic experiences*

3.2.2.3.1 Scenario Description

This scenario was built on data analysed from the profiles and preferences of the DataVaults app users, in order to create services that target tourists and citizens visiting the city of Piraeus. During this scenario, the data analysed would be used to generate reports that will assist the departments of the municipality to better design their strategies regarding the services offered to meet the touristic and cultural event demand.

3.2.2.3.2 Validity of Scenario

The Scenario remained valid, as initially described. The collected number of datasets was somehow less than initially expected, due to the limited time in running the final phase of the demonstrator and therefore openly engaging participants. The results coming out of the analysis have been fed to the local Destination Management Organization to be taken into consideration during the selection of events for the upcoming Piraeus Sea Days festival.

3.2.2.3.3 Objectives of PIRAEUS Scenario C

Objective	Status	Phase Achieved
Setup the DataVaults PersonalApp to contain the necessary for this scenario information	Achieved	Final
Extract lists of citizens from existing apps (e.g., PireApp or the upcoming touristic application) to be used for personalized cultural and touristic experiences	Achieved	Beta
Brand the DataVaults PersonalApp, to be able to promote it through the Municipality of Piraeus channels	Postponed	See D7.4
Promote DataVaults PersonalApp to citizens (e.g., via e-mail or through the website) and tourists (e.g., via posters and leaflets at the Cruise Terminal)	Achieved	Final
Advertise and implement compensation (e.g., perks) for connecting the DataVaults app with data sources (based on the features that will be offered by the platform	Achieved	Final
Push questionnaires to citizens and tourists in order to extract information in the absence of other data sources	Achieved	Final

Table 10: PIRAEUS Scenario C objectives

3.2.2.3.4 Scenario C: Evolution during Final phase

The initial step for Scenario C was to collect citizen's information and form a basic group of users, using data from Pireapp, the Municipal app used to collect citizen's complaints, ideas and suggestions, also used as one of the main channels for informing citizens about what

happens in Piraeus. Using Pireapp information we have reached out to citizens and tourists, promoted DataVaults and managed to form a final group of over 150 users for our pilots.

The second step was to demonstrate to all participants the use of the DataVaults PersonalApp. Several sessions were conducted in order to demonstrate its use to tourists. The participants in these sessions were asked to expect the questionnaires that the Piraeus pilot was planning to deploy in order to obtain the required personal data.

An extra effort to promote DataVaults app and engage more tourists was given for this scenario, as a direct short demonstration of the app was held at the cruise port of Piraeus. Cruise passengers were engaged when disembarking and waiting for transportation to the city's points of interest. The DataVaults app was showcased through a tablet and people were asked to take part in the Piraeus pilot.

Questionnaires

The next step for us was to develop these questionnaires, based on the information and the experience we had from the paper questionnaires previously deployed to establish baselines KPI values, under the guidance of the Piraeus Destination Management Organization.

The aforementioned questionnaires were created within the DataVaults app, finalized and deployed to the Piraeus pilot participants. It consisted of 7 questions, 2 demographical and 5 topics related. A value of 50 points was given for answering the questionnaire. At the time of downloading the answers in order to perform our analysis, 135 people had answered the Scenario C questionnaire.

Analysis

The amount of data collected through the use of questionnaires underwent analysis using the SEAS toolbox and spreadsheets. The results were presented to the Destination Management Organization and have been taken into consideration for the selection of events in the upcoming Sea Days festival (more than 100 events taking place in Piraeus between the 27 May and 15 June). At the time of this report the events have not been finalised or publicised. More on the Piraeus Sea Days festival and the events here: https://www.imeresthalassas.gr/
.

3.2.3 High Level Impact Recorded in the Demonstrator

3.2.3.1.1 Benefits for the Data Seekers recorded from the Scenario.

As the sole Data Seeker in all three scenarios, we have had the opportunity to explore the potential benefits of the DataVaults platform for Municipal operations. We have identified common benefits for all three scenarios. The ease and speed in collecting information with respect to the current ways of gathering data is unmatched. Moreover, the ability to obtain personal information in a GDPR compliant way, without tiresome procedures and worries about the legal aspects of the data acquiring process is a strong reason for using the platform. The ability to describe the exact type of required information is also a very important point for us. In contrast, in the already available to us data sources we must extract the required set of information and in many cases, we receive data close but not exactly as required for

our analysis. Finally, the accuracy of information in the obtained datasets, especially when gathered from an automated source (e.g., routes from the DataVaults mobile app), is a major advantage of the platform for any data seeker.

The short time of running the demonstrator scenarios in the final phase with a complete set of functionalities for the platform and the mobile app deprives us from the opportunity to present a significant impact on our operations as a Data Seeker. However, the aforementioned benefits combined with the results and analysis of the obtained datasets provides a strong indication that the DataVaults platform is an extremely useful tool for any data seeker. We also believe that the potential incorporation of more data sources, especially the ones with higher use and impact, will further benefit the data seekers and will be a crucial step for the adoption of the platform.

3.2.3.1.2 Benefits for the Individuals recorded from the Scenario.

The citizens/fans/tourists, through the information collected in our scenarios are receiving:

- improved mobility services
- personalized commercial information and offers.
- personalized tourist/cultural/entertainment suggestions

Moreover, the individuals participating in the above scenarios through sharing their personal data also have the personal satisfaction of participating in the analysis and decision making process that addresses the problems in the city. Finally, the individuals are getting compensation (for our pilots tickets to the Municipal Theatre) for sharing the data.

3.2.3.1.3 Benefits for the Organisation recorded from the Scenario.

The Municipality of Piraeus has benefited from the DataVaults platform in several aspects.

<u>Scenario A</u>

For this scenario, a traffic heat map at times of sport events, was created based on the information gathered from the platform, in order for the main routes and parking locations to be identified.

Figure 14: Traffic heat map at times of sport events, as derived from DataVaults.

In the following figure derived from the Piraeus Sustainable Urban Mobility Plan (SUMP) we can see a noise heat-map attributed to general vehicle movement in Piraeus.

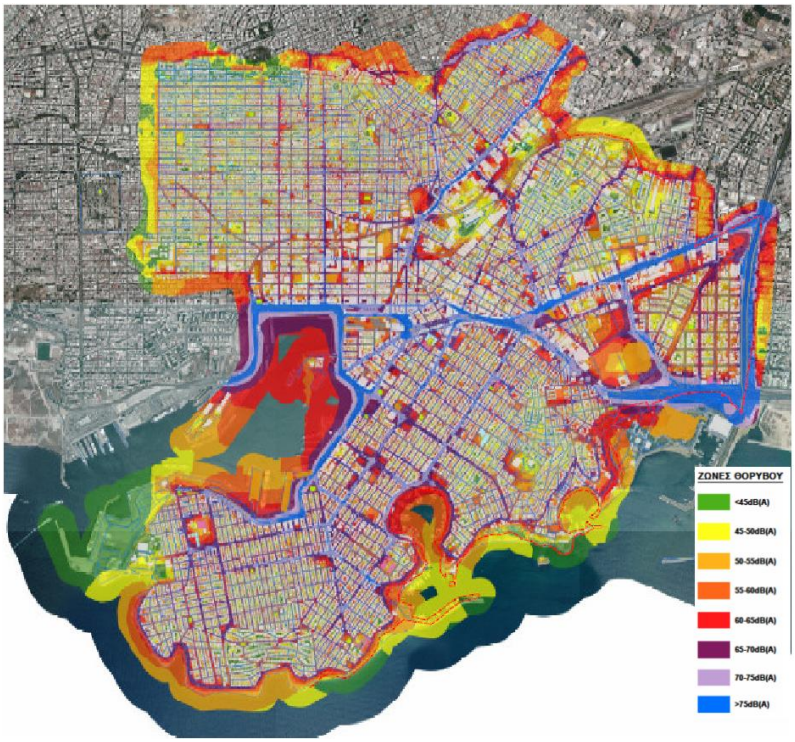

Figure 15: Noise heat map attributed to vehicle movement from Piraeus SUMP

It is evident that the two figures are complementing. In both figures the area around the stadium and the busiest routes are clearly identified. The analysis of the information derived from DataVaults has thus come to verify the findings of the Piraeus SUMP.

A more accurate analysis leading to decisions on rerouting and parking accommodation for the people also exists within the Piraeus SUMP (e.g., two different parking structures in levels around the Karaiskaki stadium, as seen in the figure below). This is also in line with our findings.

Figure 16: Proposed parking spots from Piraeus SUMP

More info on the Piraeus Sustainable Mobility Plan can be found here: https://sump.piraeus.gov.gr/ .

Finally, the Piraeus Municipal Police was presented with the aforementioned analysis and is planning a small presence of officers on the adjacent road east of the stadium, in order to ease traffic loads through the route at times of events.

Scenario B

For this Scenario, a simple analysis has led to useful results in terms of consumer preferences for the Piraeus market.

Clothing, footwear and consumer electronics showed an increased interest.

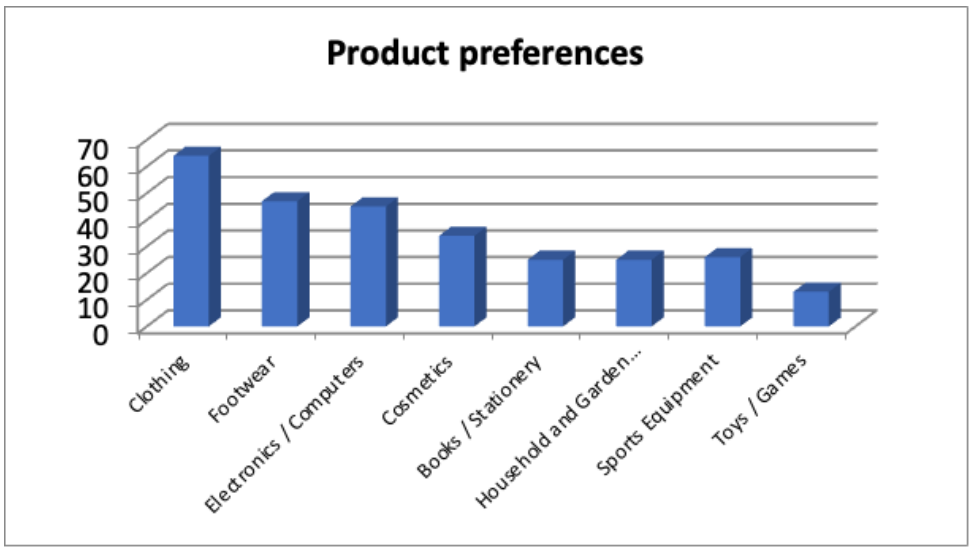

Figure 17: Consumer preferences as derived from DataVaults questionnaires.

Another important piece of information comes from the purchases outside the Piraeus market. In the next figure we can see that an even distribution between products exists, with a higher preference in footwear and electronics. The Piraeus Trade Association has been presented with these results and is intending to engage in promotion of the local market during the Piraeus Sea Days festival, which attracts a lot of people in the city.

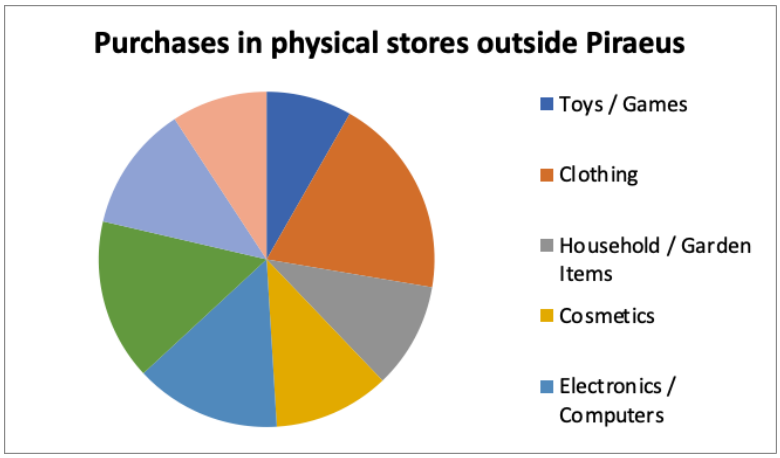

Figure 18: Purchases in stores outside Piraeus market, as derived from DataVaults questionnaires.

Scenario C

Like the previous scenario, in scenario C we have analysed the data shared from the deployed questionnaires. A strong preference in food/restaurants, theatre/cinema and live music/concerts was extracted from the analysis, as shown in the following figure 19. Based on this the Piraeus Destination Management Organization is planning specific events during the 2023 Sea Days festival, to fulfil the desires. Moreover, the extracted from the questionnaires preference on Greek and European cuisine (figure 20) is taken into account for the details of an open kitchen event that is being planned to be part of the upcoming Sea Days festival.

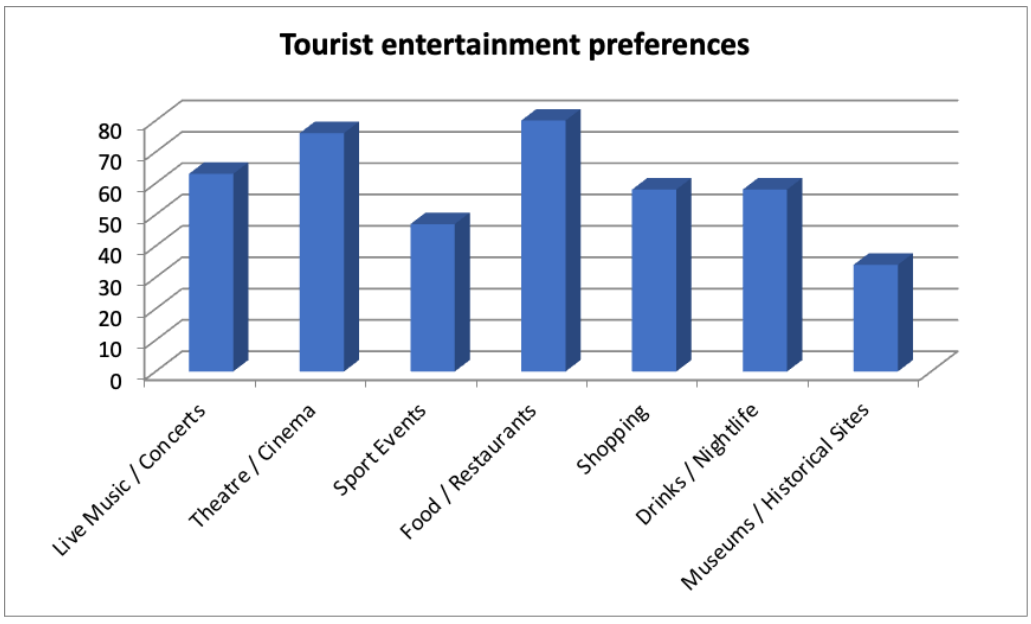

Figure 19: Tourist entertainment preferences, as derived from DataVaults questionnaires.

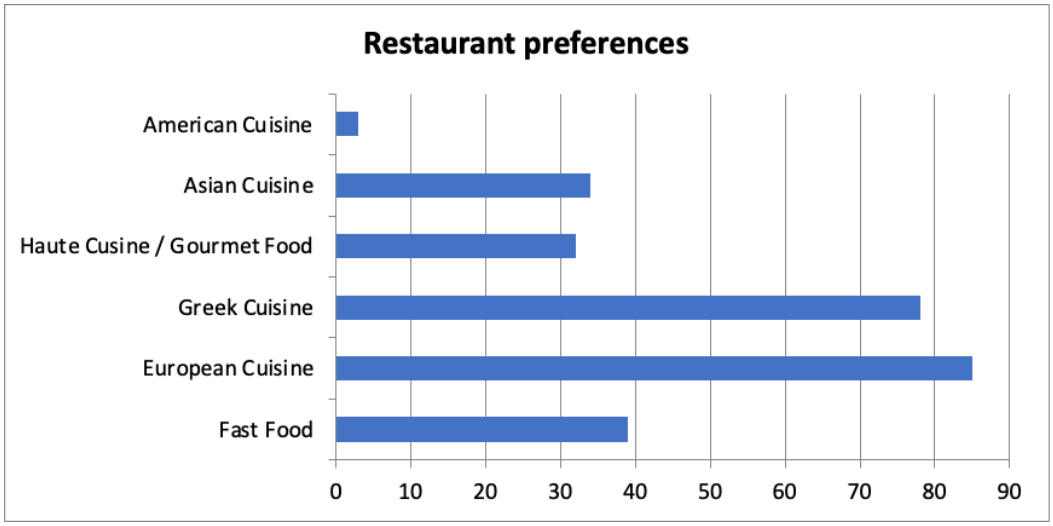

Figure 20: Restaurant preferences, as derived from DataVaults questionnaires.

3.2.3.1.4 Demonstrator KPIs

In this section, we present the KPIs relevant to the demonstrator as identified in deliverables D6.1 and D6.2, as well as other KPIs recorded during the demonstration phase.

It is noted, that not all these KPIs could be measured during the demonstrator phase of the project, as some of them concern post-project targets, as part of the exploitation activities of the demonstrator. Those are marked as such "Long-Term Impact" KPIs and are discussed in deliverable D7.4. Moreover, several KPIs were affected by COVID or other ongoing efforts in the city of Piraeus (e.g., number of citizens actively sharing data is increased also by promotion of other Piraeus apps, touristic activity is highly increased by the fact that Piraeus has become a cruise home port, number of customers entering stores and commercial activity has risen due to the metro and tram opening). The short time that the demonstrator operated with a fully operational platform and app has slightly contributed in most of the KPIs. However, we are confident that the potential to reach the desired values when and if DataVaults is adopted by more end users is clear.

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Status at M40	Comments
To increase the number of citizens actively sharing data.	Currently 500 citizens	4,000 citizens	Registered Users	7271	Registered users in Pireapp mobile app and Open1 web portal.
New municipal services.	N/A	3 new services for the municipality	Information by the Local Authority	4	Municipal Police presence during sport events, 3 Sea Days events from DMO

Improved citizen's satisfaction with services.	N/A	60% increase in satisfaction levels.	Surveys	26%	As derived from baseline and final questionnaires.
Touristic activity.	16.000 tourists	27.000 tourists per annum.	Number of visitors reported to the local authority.	880.416	Number of cruise ship visitor as reported by the Piraeus Port Authority.
Improved local commercial activity.	€1.200.000 as current	Increasing to €2.000.000	Turnover of local retail as reported to local authority.	Over 8.000.000€	Municipality estimation based on reported (for taxation) by the local shops turnover.
Increase in the number of customers entering the local stores	Increase in the number of customers entering the local stores as percentage	Not available	10% increase	23%	As derived from baseline and final questionnaires.
Increase in revenues of the local shops participating in the pilot	Increase in revenues of the local shops participating in the pilot as percentage	Not available	10% increase	18%	As derived from baseline and final questionnaires.
Number of entrepreneurs involved	To be extracted from the platform	0	20	22	Small group of entrepreneurs was formed during the alpha and beta phase.
Number of shared datasets	To be extracted from the platform	0	50	340	316 questionnaires and 34 routes
Decrease in tie required to reach the sports venue	Decrease in time required to reach the sports venue as percentage	Not available	10% decrease	1%	As derived from baseline and final questionnaires.
Decrease in time to park around the sports venue	Decrease in time to park around the sports venue as percentage	Not available	10% decrease	1%	As derived from baseline and final questionnaires.
Number of tourists and citizens participating	To be extracted from the platform	0	200	150	

Number of data analysis procedures	Number of data analysis procedures run by the local Destination Management Organization	0	3	2	
Number of actions taken by the local Destination Management Organization based on Datavaults data	Number of actions taken by the local Destination Management Organization based on DataVaults data	0	5	0	Postponed as pilot was run late in the program. 3 events are planned for the upcoming Sea Days Festival.

Table 11: Demonstrator #2 KPIs

3.2.4 Demonstrator's Activities Timeline

The following table presents the main activities performed within the demonstrator.

Demonstrator 2 PIRAEUS	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40
Scenario A - Smart Mobility Services for Individuals																						
Alpha Phase																						
Collection of data																						
Examine and analyse data																						
Beta phase																						
Open up the use of DataVaults personal app to greater groups of users																						
Acquire analytics																						
Final Phase																						
Make the branded version of the DataVaults personal app available to everyone																						
Acquire advanced analytics from the DataVaults cloud platform and propose actions for smart mobility																						
Activate the sharing compensation mechanisms																						
Scenario B - Empowering local entrepreneurship																						
Alpha Phase																						
Collection of data																						
Examine and analyse this initial data																						
Beta phase																						
Open up the use of DataVaults personal app to greater groups of users																						
Acquire analytics																						
Final phase																						
Make the branded version of the DataVaults personal app available to everyone																						
Acquire advanced analytics from the DataVaults cloud platform and propose actions to empower local entrepreneurship																						
Activate the sharing compensation mechanisms																						
Scenario C - Services for Personalized cultural and touristic experiences																						
Alpha Phase																						
Collection of data																						
Examine and analyse this initial data																						
Beta phase																						
Open up the use of DataVaults personal app to greater groups of users																						
Acquire analytics																						
Final phase																						
Make the branded version of the DataVaults personal app available to everyone																						
Acquire advanced analytics from the DataVaults cloud platform and propose personalized cultural and touristic experiences																						
Activate the sharing compensation mechanisms																						

Table 12: Execution Timeline for Demonstrator #2– PIRAEUS

3.3 DEMONSTRATOR #3 - SECURE HEALTHCARE DATA RETENTION AND SHARING

Andaman7 is a patient centric mobile tool for collaborative management of medical records between health professionals, patients, authorized relatives of such patients (parents or spouse, for example) and connected devices. Exchange and synchronization of information are Peer to Peer, based on original concepts maximizing security and confidentiality.

The main aim of the A7 demonstrator was to integrate DataVaults into the A7 mobile application to showcase how this integration can be performed and to add DataVaults as a repository to collect and store their A7 data, and as a new source for allowing A7 users to collect data coming from other sources as well.

3.3.1 Target Audience Reached during Final phase.

3.3.1.1 *Data Owners*

As planned, few Andaman7 users were reached to participate in the demonstration. The recruitment process was followed but not as strictly as first intended. Given the timeframe left to complete this process, some shortcuts were taken to be able to collect enough data and produce good outcomes.

A first limited group of French speaking users was contacted - corresponding to users who answered the first citizen questionnaire distributed during year 1 but none were interested in testing the application. We then decided to extend the reach to all Andaman7 French users. From those, 66 users show interest to enter demonstration.

As the number of interest users is rather short, inclusion/exclusion criteria were applied in a very light manner. From those users, 56 were invited to join demonstration activities. At this time, 20 users managed to create a DataVaults account and connect Andaman7 to the platform for at least one of the two parts of the scenario A.

3.3.1.2 *Data Seekers*

For now, no data seekers were directly reached but a few newsletters were sent concerning the DataVaults projects from the beginning of the project and many of our data seekers contacts have received some information about the project and future activity is reported in D7.4

Our main data seekers target remains companies running clinical trials as we believe that the combination of Andaman7, as a collector, and DataVaults, as an aggregator brings lots of potential in such kind application. A white paper listing all functionalities this combination can bring to clinical trials was then written.[8] This document will be communicated to several contacts in this field at the end of the project. Sadly, this will be too late to be able to collect and document feedback in this document, but health is a difficult field and reach such companies with an incomplete document would have discredited the project.

[8] https://drive.google.com/file/d/1ViWLEuSaBBhnZBLMXIT7012c_1OBbXJc/view?usp=share_link

3.3.2 ANDAMAN7 Demonstrator Scenarios Evolution

3.3.2.1 *ANDAMAN7 Scenario A - Get data from Andaman7 and provide to health sector.*

3.3.2.1.1 Scenario Description

In this scenario, current users of Andaman7 will be able to connect to DataVaults to store all or part of their health data. This storage can be used as a backup to retrieve data when lost. This can also be used by third parties in the health sector (e.g.: clinical trial, research). Due to some delay that prevented the starting of demonstration activities with final users earlier, we had to do some cuts to reach our goals. The quantity and variety of users involved and collected data is a bit reduced from our first expectation, but we still managed to reach our goals by also involving more Andaman7 employees in the project during those last few months. The resulting analysis of the demonstration activities can then be considered as unchanged even though the quality and quantity of collected data would have probably provided more meaningful results.

3.3.2.1.2 Objectives of ANDAMAN7 Scenario A

Objective	Status	Phase Achieved
Develop the backup of Andaman7 content through the DataVaults platform	Achieved	Alpha
Develop a connector to upload Andaman7 data to the DataVaults platform	Achieved	Alpha
Develop and run a fictitious clinical exercise	Achieved	Final
Patient recruitment and study recruitment	Achieved	Final
Service creation to integrate onboarding process into Andaman7	Achieved	Beta
Development of a connector to download DataVaults data and integrate it in Andaman7	Achieved	Beta
User Onboarding	Achieved	Final
Data collection	Achieved	Final

Table 13: ANDAMAN7 Scenario A: objectives

3.3.2.1.3 ANDAMAN7 Scenario A Evolution

The actions taken for the realisation of the scenario were the following:

- Throughout the whole project, the DataVaults platform was tested, and feedback was provided to ensure its reliability and efficiency.
- Andaman7 connectors were developed and tested to enable the automatic upload of Andaman7 data to the DataVaults PersonalApp. From those connector, new features were developed and added to the Andaman7 app, notably the backup feature:
    - The Andaman7 app was equipped with backup components to ensure that the data was securely uploaded and stored to the DataVaults platform and easily retrievable from the PersonalApp to allow restoration within Andaman7. All

those components were deeply tested to ensure a proper solution to secure health data of the final users.

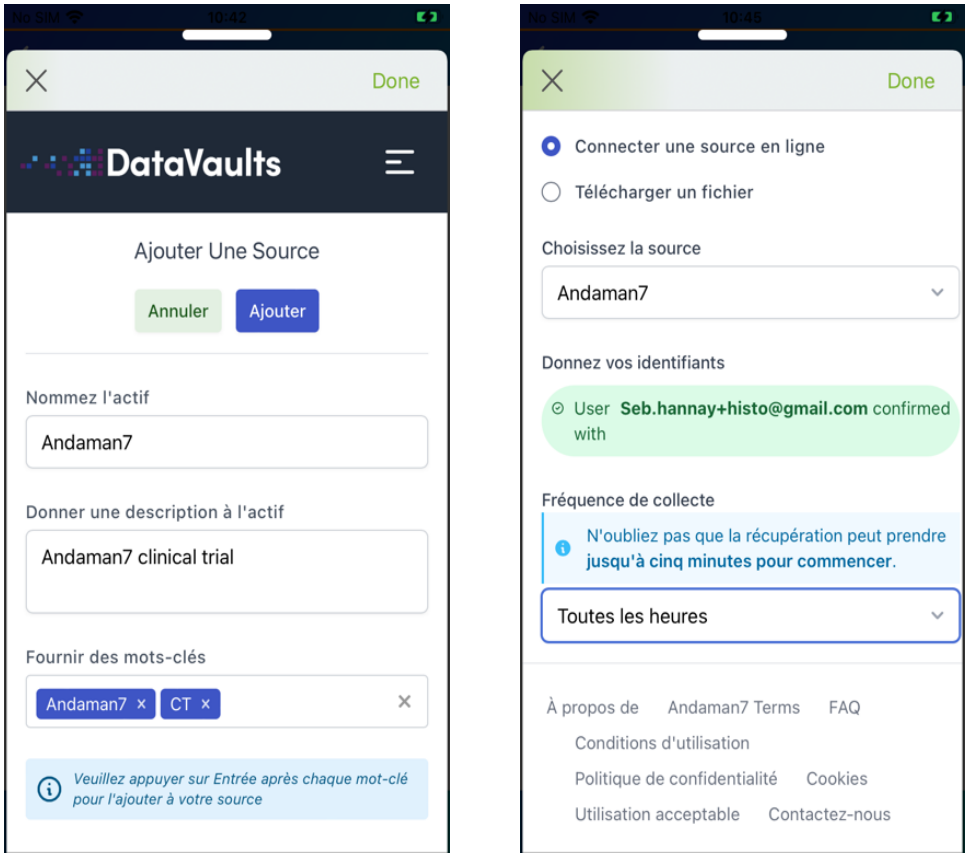

Figure 21: Configuration of Andaman7 data source connector from A7 app

- The Andaman7 app was updated with four new services to integrate those new features. Those services activate them and onboard users into the specific configuration of each scenario directly within the Andaman7 app. The onboarding process configured to allow patients to choose what they want to share with the DataVaults platform to provide maximum control on data.
- Patients were recruited and onboarded onto the platform, and data was collected from them through the Andaman7 app and securely stored on the DataVaults platform. Information sessions and webinars were conducted to educate patients about the platform and its benefits.

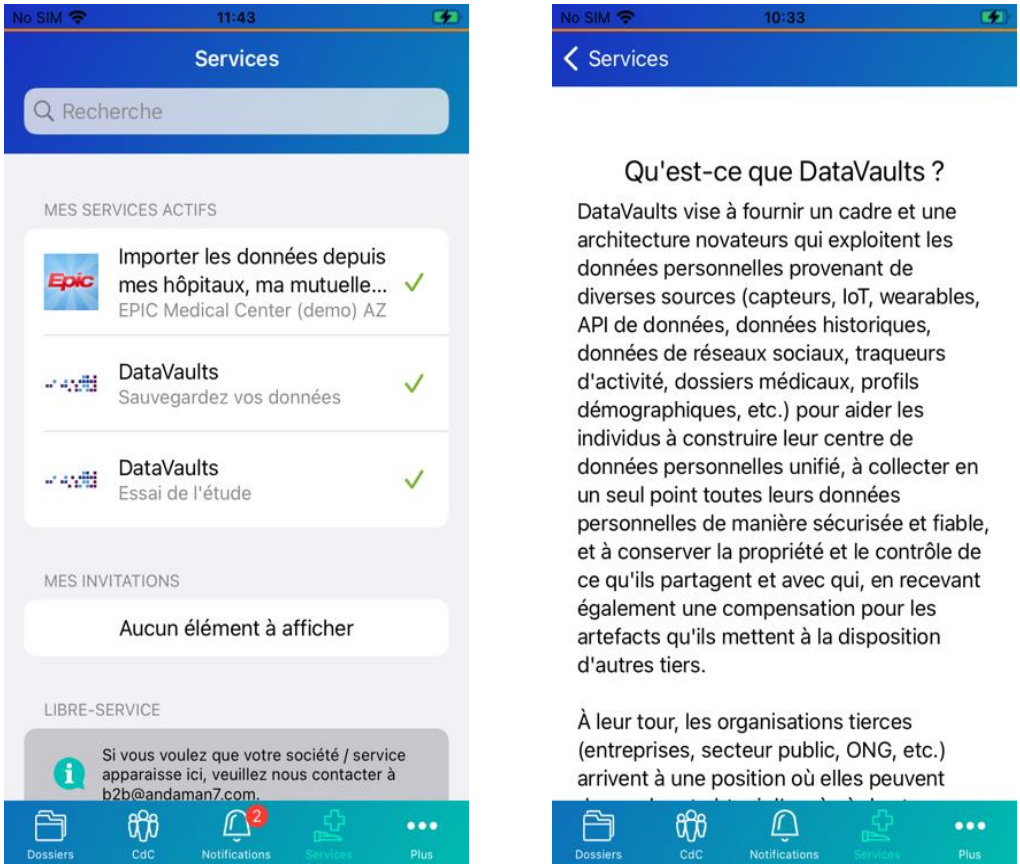

Figure 22: Onboarding process integrated in A7.

- Patients were recruited for clinical trials and onboarded onto the platform. Data was collected from them using the Andaman7 app and securely stored on the DataVaults platform. Andaman7 team simulated a data seeker and retrieved data through the DataVaults cloud platform. The data was then analysed to draw insights and conclusions on what can be done in such field.

Data sources

For this scenario, the only data source was the electronic health record stored in the Andaman7 application. Those are raw health data (Andaman7 proprietary format), documents and standardised data that are already collected from many sources (some hospitals, connected devices, other Andaman7 applications, etc.). It can be divided in five main categories: medical data, health data, fitness data, sleep data and nutrition data.

The second part of the scenario (clinical trials) included an additional category of data: Questionnaires. Those questionnaires are directly added into the health record and then have the same storage format as any other Andaman7 data. To avoid extended ethics research during the project, a fictitious clinical study was created. This trial is based on our knowledge in this field. The data collected during this demonstration study was retrieved through the DataVaults cloud platform. We then explored all the possibilities DataVaults offers to analyse those pieces of data. All findings of this analysis are available in a clinical trial white paper that will be used to communicate with potential data seeker in this field.

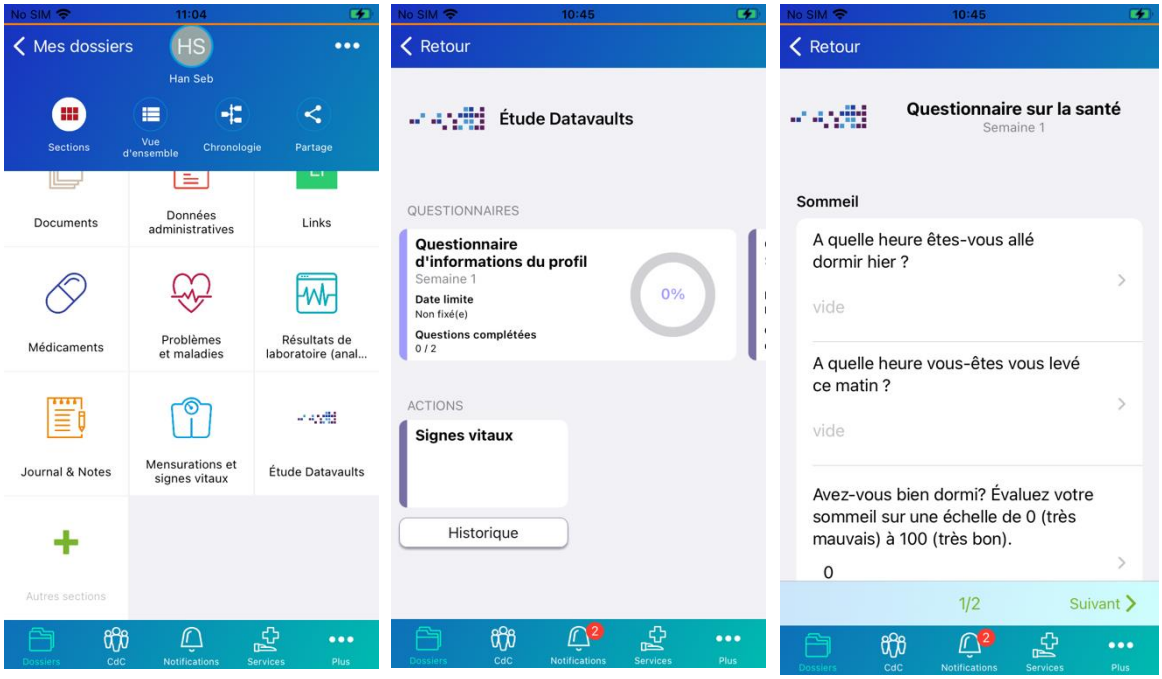

Figure 23: A7 Clinical trial integration

Scenario Findings

The collected feedbacks show that configuring the DataVaults platform and the way to connect the Andaman7 app to it is not a simple task. This connection is not as integrated as we were expecting at the beginning of the project. Our first expectations were to be able to connect through API calls and so the whole connection process would be integrated directly in the Andaman7 app without showing the DataVaults app. This was not possible to realise during the allocated time.

We fallback to a semi-integrated version which is working perfectly but needs more step from the user as he needs to configure the sharing from Andaman7 then do the connection within the PersonalApp through an integrated web view. To facilitate as much as possible this connection, a few onboarding/demo sessions were run to educate the user. Despite those sessions, the whole process was not easy to follow and understand for some part of our userbase, which count many older people which are not used to technology in general.

In addition, the whole DataVaults platform, the way of displaying data, the way to configure sharing to data seekers was reported as difficult to understand. Again, the fact that most of our user are not used to such kind of technology can explain but in a general manner this was to be expected. The DataVaults platform allows great security and control on data. Even though it brings some pre-configuration that ease the process, it's not easy for an older user to really understand what he does.

Of course, that kind of configurations is needed when you are dealing with sensitive personal data like health data, but this also means that, most probably, only some advanced users will be able to really understand what they are sharing. The target should then be focus on younger users, that are more used to such kind of data sharing. But these observations will be used to improve the DataVaults design.

From the medical research perspective, Andaman7 already provides an elegant and pretty complete solution to run decentralized clinical trials. But Andaman7 doesn't propose the whole stack of services related to medical research.

Andaman7 offers the following main phases of clinical trials, and more generally of studies:

Figure 24 A7 Focus Area

This is where the addition of the DataVaults platform is helpful. It can improve some of the features already done by Andaman7, for example the data collection which is proposed in a new way with its own specificities. But it can also cover the two main missing phases of the stack. Detailed findings about medical research are described in the Andaman7 white paper for medical research referred to above.

3.3.2.2 *ANDAMAN7 Scenario B - Data collection and Andaman7 improvements*

3.3.2.2.1 Scenario Description

In this scenario, current users of Andaman7 will be able to connect to DataVaults to collect their health data coming from new sources. This will make the data available to patients for reviewing, learning, using in other set ups (e.g., share additional data with their doctors, hospital, etc.). Data will mostly be raw personal data but also aggregated data (based on data collected in scenario A). As stated earlier, delays encountered in data collection by the DataVaults platform lowered our initial expectations in the quantity of data that we would be able to collect by the end of the project.

Effort was focused on the realisation of proof-of-work integration (POW) based on "easy-to-retrieve" raw data. Data will be included manually in the DataVaults platform through one of the supported means at the final stage of the platform (mostly through import of CSV and PDF file directly to the platform). Also, basic analytics on collected data were executed at a high level.

3.3.2.2.2 Objectives of ANDAMAN7 Scenario B

Objective	Status	Phase Achieved
Develop a connector to the DataVaults platform	Achieved	Alpha

Develop data integration	Achieved (as POW)	Final
Patient recruitment	Achieved	Beta
Service creation to integrate onboarding process into Andaman7	Achieved	Beta
User onboarding	Achieved	Final
Involve new sources	Achieved (as POW)	Final
Analytics	Achieved (as POW)	Final

Table 14: ANDAMAN7 Scenario B objectives

3.3.2.2.3 ANDAMAN7 Scenario B Evolution

The following is a description of the actions taken for the realization of the scenario:

An analysis was conducted to identify new sources that could be integrated into the system based on the time left for the project. We finally decided to focus only on quick wins as there were delays in properly integrating all the connectors on the DataVaults platform: CSV, PDF, etc.

New Andaman7 services were developed to integrate onboarding of users for these new features. The patient recruitment and onboarding process was then initiated to begin the data collection process. This was done to ensure that the system had sufficient data to operate effectively. The data collected was stored in DataVaults.

Finally, a high-level data analysis was conducted to gain insights into the collected data. The analysis was performed to identify patterns and trends that could be used to improve the system's functionality. Ways to integrate those data into Andaman7 were also analysed and tested internally but not provided to the final user as the final solution was not mature enough to risk any potential data pollution.

Data sources

In addition to the data sources referenced in scenario A, this scenario involves new data sources with different data formats. From those, we can identify documents, mainly on the form of PDFs, which are then integrated as document in Andaman7 and CSV which can be parsed to extract raw information and integrate it as new Andaman7 data.

Besides, we analysed some calculation that can be done on Andaman7 collected data to produce potential new data that can be easily integrated.

Scenario Findings

As expected, integrate new sources is something hard to do. Each source requires a specific development to be integrated in DataVaults, which takes time and resources. In addition, if we want to integrate such data back into Andaman7, it also requires a specific development for each data source as none provides a common standard interface to retrieve them (or are limited to one single field and not well established for now in Europe). It then takes time and effort for each piece of data.

Producing analytics based on Andaman7 data is easier as they share a single data format and analytics can then be easily produced in that specific format. But calculation on health data

should always be taken with great care. If anything is wrongly calculated, it can indicate bad information to the user, which may have consequences. Even if calculations are correct, some can be misunderstood or interpreted in a wrong way, which can lead to bad responses or behaviour by the patient.

3.3.3 High Level Impact Recorded in the Demonstrator

3.3.3.1.1 Benefits for the Data Seekers recorded from the Scenario.

The main benefits recorded by the demonstrator were the following:

- Ease and lower price of some existing activities: As an example, during clinical trials, a huge budget can be allowed just to be able to find patients that match certain criteria. The only way to reach them is by contacting hospitals that will, upon patient consent, initiate the contact. A few solutions emerged recently that allows to reach multiple hospitals at once in an automated way and so reduce this searching cost. But in all of them, the hospital is used as a third party. With DataVaults, the patient is in control, which should reduce the cost.
- All-in-one solution with Andaman7 and DataVaults: Andaman7 act as a data collector. It can collect different kinds of data (medical, health, fitness, sleep, nutrition) collected from different sources (users, doctors, IoT, …) but also run specific and customizable questionnaires while staying at home (continuity of care, care path, …). Data collected can easily be retrieved and aggregated by the DataVaults platform.
- Analytics: In addition to the collection of anonymous data, tools integrated in the Andaman7 platform allow to run some analytics in them. Those analytics can be simple generic calculation but also fully personalised depending on the needs of each clinical trial.

3.3.3.1.2 Benefits for the Individuals recorded from the Scenario.

The benefits encountered by the individuals that have used DataVaults, as recorded in this demonstrator, are the following:

- New (awaited) features to Andaman7: The DataVaults platform brings new capabilities to Andaman7.
    - Backup all my medical data to the PersonalApp: Andaman7 is based on a decentralised system that brings multiple advantages but doesn't save securely data in the cloud. DataVaults fill this missing spot by provided a secure cloud-based automatic backup of all medical data stored on Andaman7. At any time, the user can retrieve those data from the personal cloud and reintegrate them in Andaman7. This is especially helpful when you lose or break your phone.
    - Get data from new sources: the DataVaults platform brings new connector that are not integrated in Andaman7. Those data can be directly imported in your PersonalApp as a separate dataset, but you can also extract it and integrate it into your Andaman7 and complete even more your medical record.
- One place to securely store all data: DataVaults brings a secure and reliable vault to store your medical data on the cloud.

- Be in control of own data: once available on the PersonalApp, dataset can be shared securely with advanced setup (anonymisation, target, …) that allows a perfect control of your data. The data owner can get rewards by sharing its data.
- Future improvements: Such kind of integration opens ideas for new improvements that could be brought to the Andaman7 platform in the future and will bring even more benefits to the user.

3.3.3.1.3 Benefits for the Organisation recorded from the Scenario.

For Andaman7, the addition of DataVaults platform brings an added value to our existing platform. From a B2C perspective, it gives new possibilities and opens the door to brand new functionalities awaited by our end-users. From a B2B perspective, it allows us to improve our current solution by simplifying some of our paid functionalities and to add new ones. All this brings more attractiveness to our existing platform and should bring new users and clients.

Our KPIs were mainly focused on increasing this added value to convince both final users and potential partners. The goals to add new features and new datatypes is achieved even if they still need to be improved for production usage. The rest was focused on long term KPIs resulting from those added elements.

Of course, what was developed during those demonstration activities is just a taste of what DataVaults can bring to Andaman7 and still needs to be improved to be used in a real-world environment as collected feedback shows. However, our end-users show the interest on what we can bring by developing more of such a solution.

3.3.3.1.4 Demonstrator KPIs

In this section, we present the KPIs relevant to the demonstrator as identified in deliverables D6.1 and D6.2, as well as other KPIs recorded during the demonstration phase.

It is noted, that not all these KPIs could be measured during the demonstrator phase of the project, as some of them concern post-project targets, as part of the exploitation activities of the demonstrator. Those are marked as such “Long-Term Impact” KPIs and are discussed in deliverable D7.4

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Status at M40	Comments
To increase active users operating Andaman7 *Scenarios: a and b*	22,000 registered users at M1 1,200 Average monthly users (2019 average)	An increase to 40,000 by M40 An increase of average registrations/day An increase of average active monthly users to 1,600	Andaman will continue to monitor these metrics using its database and Firebase Analytics tool.	Registered users: 38240 Average active users: 1,200	

To increase the number of datatypes used by Andaman7. *Scenario: b*	Current use of 100 data types supported by the app. Implemented data type in A7 app + Analytics (Firebase) on usage of each data	An increase to 120 Usage of these data types	Most will be from calculated data, analytics, charts which can be accessed through the DataVaults platform. Also Implemented data type in A7 app + Analytics (Firebase) on usage of each data	Increased to 243 to mitigate	
Increase of Volume of data per/ category *Scenarios: b*	Analytics (Firebase) of volume of data per user for each data category	To be split by categories calculated (in average 326 pieces of data per user for all categories) To show an increase of 15%	An increase should be in the number of data registered in the app thanks to new data sources. Analytics (Firebase) of number of data/users for each data categories	Increase of 29% per user (all categories) to mitigate	
Generation of new services offered through Andaman7. *Scenarios: a and b*	Currently 2 services	Goal is to have 4 1. Back-up of data 2. Share data to A7 partners* 3. Get data from new sources* 4. Get analytics data from DataVaults. *May depend on Andaman7 partners but base can be done if 100% fulfilled.	Implemented services in A7 app + Analytics (Firebase on usage of service)	4 services added	2 are fully functional and 2 still need to be improved to be ready for real world.

Table 15: Demonstrator #3 KPIs

3.3.4 Demonstrator's Activities Timeline

The following table presents the main activities performed within the demonstrator.

Demonstrator 3 ANDAMAN7	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40
Scenario A - Get data from Andaman7 and provide to health sector																						
Alpha Phase																						
Dev. upload connector																						
Dev. download connector																						
Dev. backup																						
Patient recruitment																						
Beta phase																						
Service creation																						
Users onboarding																						
Data collection																						
Fictitious clinical study																						
Study recruitment																						
Final phase																						
Study data collection																						
Result documentation																						
Data seeker presentation / onboarding																						
Scenario B - Data collection and Andaman7 improvements																						
Alpha Phase																						
Dev connector																						
Dev data integration																						
Patient recruitment																						
Final phase																						
Service creation																						
Users onboarding																						
Analytics																						
Involve new sources																						

Table 16: Execution Timeline for Demonstrator #3 – ANDAMAN7

3.4 DEMONSTRATOR #4 – SMART HOME PERSONAL ENERGY DATA (MIWENERGIA)

MIWenergia is an SME electricity retailer founded in Murcia, Spain. It operates in Spain nationwide with a varied range of customers: residential, industries, small and medium enterprises, and large tertiary buildings. Providing service to more than 3.000 customers with an annual energy transit of 100 GWh, MIWenergia offers clients additional services related to energy efficiency such as energy consulting and audits, design and installation of photovoltaic systems for self-consumption, electric vehicles charging stations, and energy monitoring.

A platform like DataVaults, where all the user generated data is provided to the platform by the data owner, can give access to all the companies interested in those data, connecting the data processors, distribution system operators, or electricity retailers, and letting other energy entities use that detailed information about electricity consumption with the user's permission. MIWenergia considers a platform for accessing data like DataVaults needed due the low quality of the platforms and data which are available now. Furthermore, the company only has access to the data of their customers, DataVaults gives the opportunity of having the data of all the platform users. MIWenergia will utilise DataVaults, to try and acquire such data directly from customers that are willing to share those, towards providing better services to them.

3.4.1 Target Audience Reached during Final phase.

The following sub-sections provide a status update on the audience reached by the demonstrators during the final phase of the demonstrator.

3.4.1.1 *Data Owners*

In the final phase of the demonstrator, the recruitment of participants was finished, with the involvement of 20 participants (15 MIWenergia's clients and 5 non-clients) in the demonstration activities of the platform. We have performed 3 demo sessions with the participants to train them on how to use the platform and show them all the functionalities. They have registered in the DataVaults PersonalApp and have tested all the functionalities available.

The MIWenergia's API and connector is fully deployed in the platform and being used by the participants.

3.4.1.2 *Data Seekers*

MIWenergia have been participating in the project not only as a bridge to the data owner providing connection to their data but also as a Data Seeker. We have tested all the functionalities available as a Data Seeker, including the Cloud and the SEAS platforms. We have contacted two more potential Data Seekers and we have shown the platform to them in two individual demo and training sessions, collecting their feedback. One of them, that represents a public entity, have decided to test the Cloud platform in order to obtain data for research purposes. The second one (a private company that is a PV installer), didn't want to be involved in the demonstration activities but it is interested in the project results regarding the analysis of energy data and the design of PV system offers (Scenario A of the Demonstrator), thus MIWenergia will keep them updated regarding this scenario results.

3.4.2 MIWENERGIA Demonstrator Scenarios Evolution

3.4.2.1 *General evolution: MIWenergia's API development*

MIWenergia as an electricity retailer manage and store energy data of their clients but also has the capability to access the data of non-clients with the proper consent from the customer. In order to provide access to the Data Owners to their electricity data in a simple way through the DataVaults platform, MIWenergia developed during the project their own API, that was integrated in the DataVaults ecosystem.

The main steps for developing the MIWenergia-DataVaults API and the MIWenergia's connector were:

1. Analysis of the available energy data
2. Design and creation of data model to manage project's users and API users.
3. Development of the API services and controllers
4. Development of the API login and securitisation
5. API development and testing
6. Writing of API documentation
7. Integration of the API in the DataVaults platform – MIWenergia connector
8. Testing and bugs fixing
9. Updates and final deployment

At this final step, the MIWenergia API is fully integrated as a MIWenergia connector in the DataVaults platform, and the Data Owners can use it to collect and store their hourly energy consumption data and share it with the Data Seekers. The MIWenergia connector can provide the real data for Data Owners between two dates selected by them and selecting the USPC (supply point) from the list of USPC available for each user.

The connector is the main channel to provide the energy consumption data to all the Scenarios of the demonstrator. The API can be accessed through the following links:

Swagger	https://api-dv.miwenergia.com/swagger/index.html
Endpoint	https://api-dv.miwenergia.com/apidv/

In the next figures there are some screenshots of the swagger and the MIWenergia connector of the DataVaults PersonalApp.

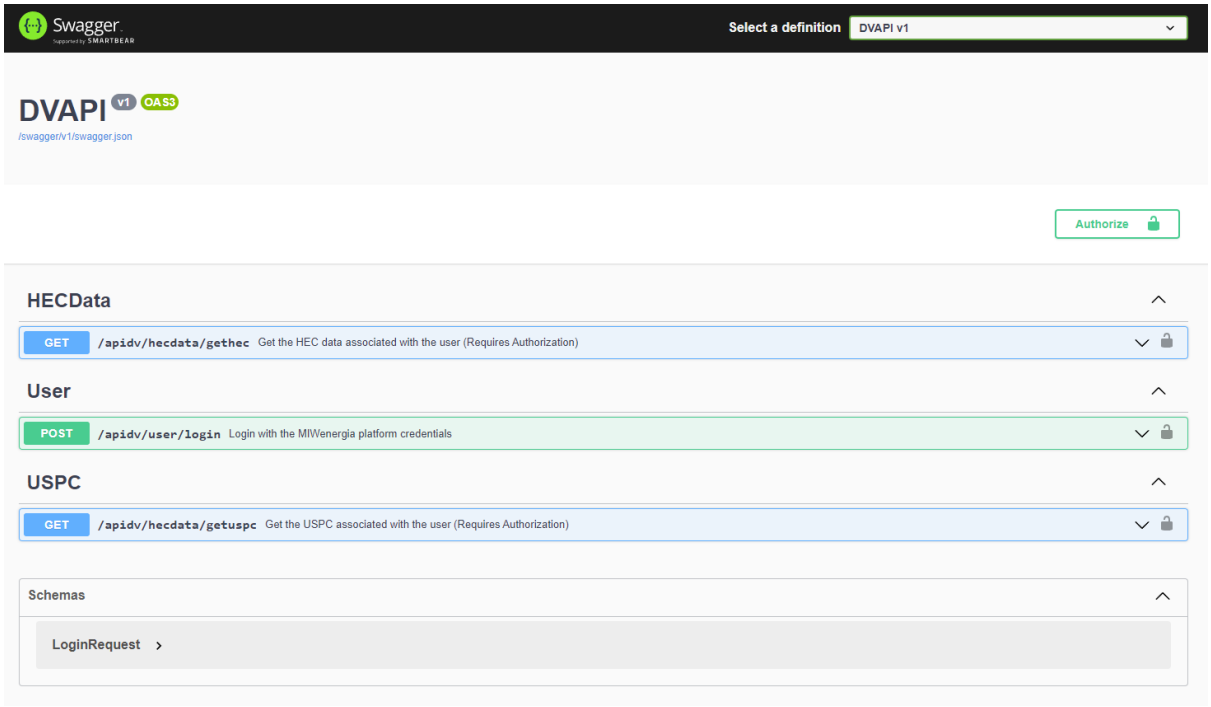

Figure 25: MIWenergia API (Swagger)

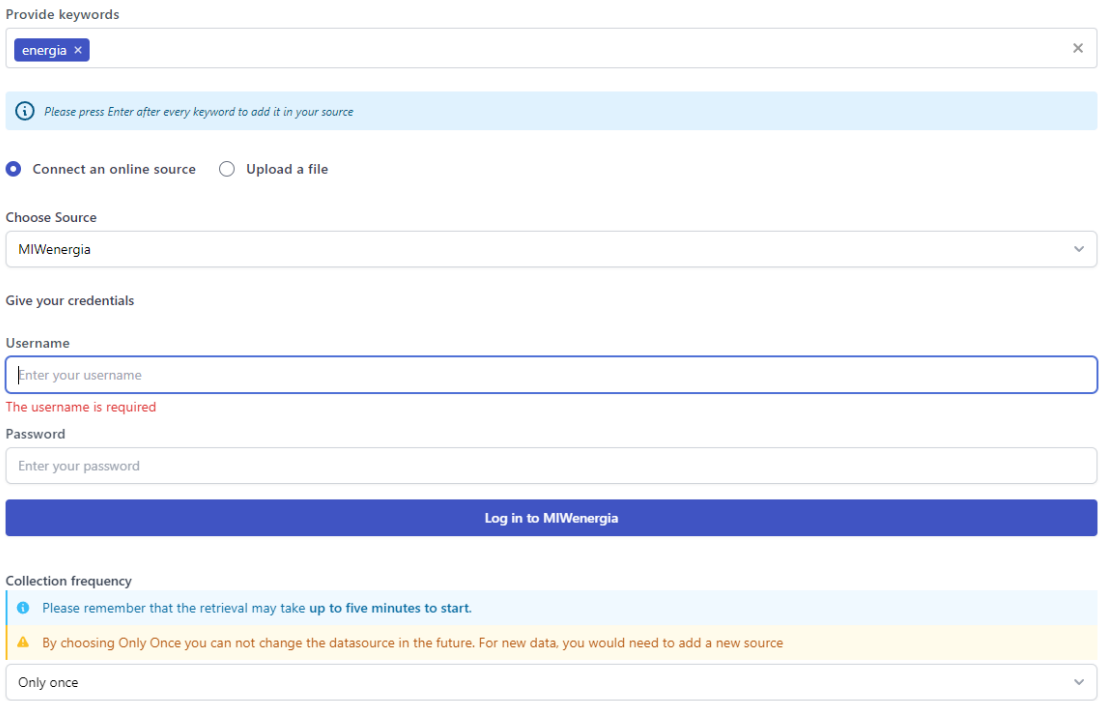

Figure 26: MIWenergia connector (I)

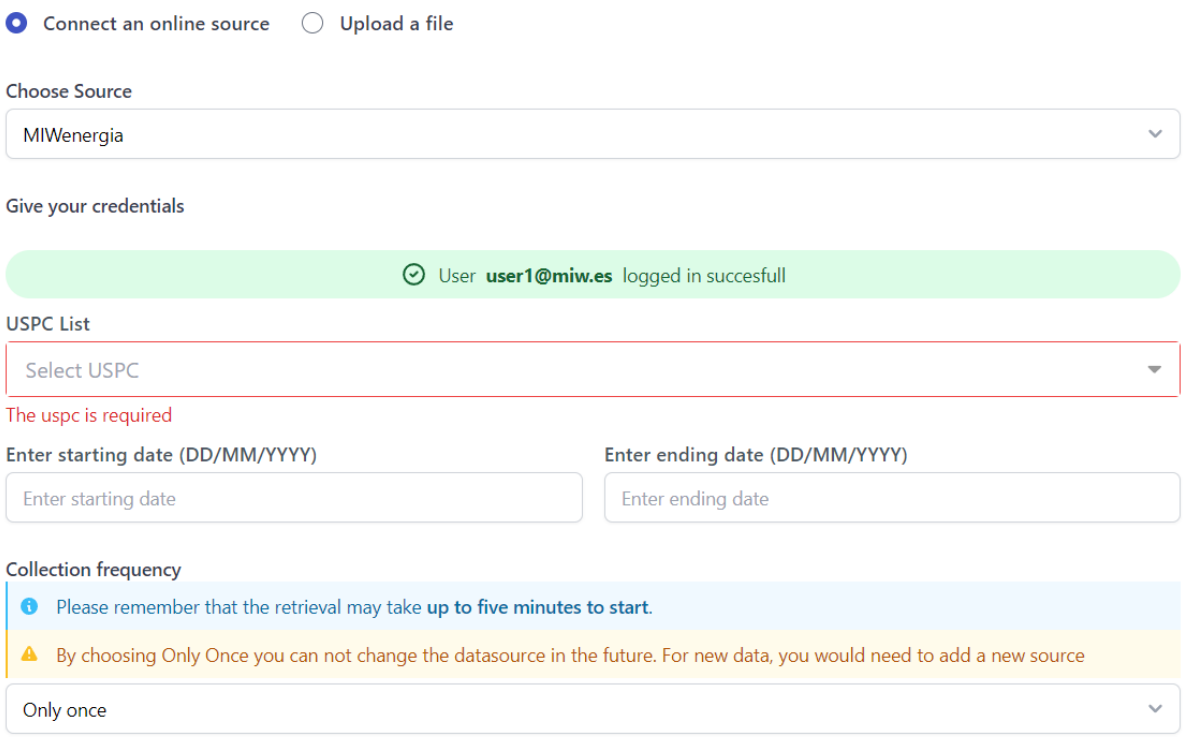

Figure 27: MIWenergia connector (II)

3.4.2.2 *MIWENERGIA Scenario A – PV installation design for self-consumption*

3.4.2.2.1 Scenario Description

The *Scenario A: "PV installation design for self-consumption"* was conceived to solve the problem of accessing hourly electricity consumption data for which the photovoltaic system needs to be efficiently designed. Normally, the PV installation companies don't have access to this data, as the Distribution System Operator (DSO) is the company in charge of this measurements. PV installers normally make estimations based on total yearly energy consumption values. Sometimes these values are real data obtained through the electricity bills but in many cases, they are only rough estimations because the clients don't have (or don't know how to) access to their own data. Some companies help the customers to access the data, but the process is long and complicated, so clients are not keen on doing so. In addition to the difficulty of accessing the data, the lack of knowledge on the energy sector, such as the units of measurement (kW, kWh, V, A) and their meaning and implications, make conversations with customers interested in PV systems not very effective, and there are frequently misunderstanding when the information is provided.

3.4.2.2.2 Validity of Scenario

The scenario remains valid although some changes have been made to adapt the approach to the final functionalities available in the platforms. As the questionnaires and the datasets cannot be related to which user has sent them, the design of the PV installation is made in a "persona" approach. There have been designed PV systems based on average profiles for the most common users in the platform and a special sales campaign for PV systems with these characteristics has been offered considering the users' average profile.

3.4.2.2.3 Objectives of MIWENERGIA Scenario A

Objective	Status	Phase Achieved
To have a functional API that returns customer's hourly consumption	Achieved	Alpha
Additional mock-ups definition	Achieved	Beta
Participant Recruitment	Achieved	Final

Table 17: MiWenergia Scenario A: objectives

3.4.2.2.4 MIWENERGIA Scenario A: Evolution Plan

The main objective of this scenario (and for the demonstrator globally) was to try to help customers and simplify the process to access, collect and store their hourly energy consumption data in a safe and private way, using an "independent" platform such as DataVaults, and give them the power to share the data with the companies (Data Seekers) they want and with fair prices. In this way, the MIWenergia connector developed during the project, contributes to solving this problem.

From the Data Seekers point of view, that in this scenario are mainly represented by PV installations companies, there is a need of having precise data for a correct sizing of a photovoltaic installation. Hourly consumption and production must be matched in order not to overestimate or underestimate the installations and make it profitable for the clients (and for the installers). In addition, using aggregated data for calculating the systems reduces the reliability of the proposals and can cause dissatisfaction among customers and lack of credibility.

Apart from the electricity consumption data, designing PV installations requires some extra information. Due to the need to collect it "manually" from users, we have designed and implemented some questionnaires in the DataVaults Cloud that have been sent through the platform to the Data Owners. In this Scenario, the minimum information required was defined:

Data	Details
Location	At least city, the more detailed the better
Available outside space	Square meters
Kind of roof	Flat or slope
Available outside space orientation	North, South, East, West

In the next Figures, it can be seen some screenshots from the questionnaire implemented from the Cloud and the PersonalApp perspectives:

Figure 28: Questionnaire from the Cloud-Data Seeker Perspective (Scenario A)

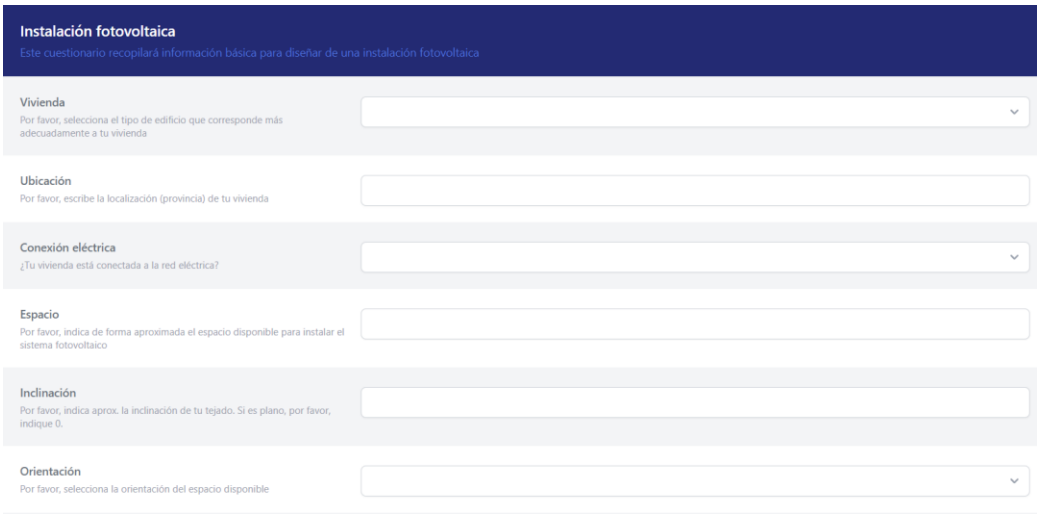

Figure 29: Questionnaire from the PersonalApp-Data Owner Perspective (Scenario A)

In parallel with the development of the API and all the DataVaults platforms, we started the recruiting process to participate in the demonstrator activities. We contacted many of our clients and non-clients that are already participating in other EU projects in order to involve them in the DataVaults demonstrator. The reason to choose clients that already participate in other projects is that they are people already interested in innovation and research project and that are aware of the dynamics and activities of this type of projects.

There were also spent great efforts for testing the platforms internally, collecting issues and bugs and giving advice from the demonstrator point of view to the technical team, helping continuously with the improvement of the platforms. In this way, we ensure that all the platforms were stable and trustworthy enough before providing access to the participants. This is a fundamental issue, as project participants lose interest easily if they feel that the

tools are not working properly. In addition, we gave participants one demo session and one training session on the project and the platform to help them understand the tools and the workflow.

Finally, once all the Platforms were ready, the data sources available, the questionnaire implemented and the participants onboard and testing the platform, MIWenergia started to act as a Data Seekers, buying, collecting, and analysing the data and the questionnaires answers obtained through DataVaults. MIWenergia have also provided compensation methods for the wallet, to motivate the participants to "play" with the PersonalApp and test all the functionalities. For the analysis of the data, the main tool used was the SEAS platform developed in the project.

Although the first approach of the Scenario was to design customized PV installations to offer to each specific customer, due to ethical and privacy reasons, the questionnaire answers and the datasets cannot be related to which user had sent it, and as a consequence, preventing Data Sekeers to "contact" the Data Owners through the platform. For that reason, we decide to modify the approach of the scenario, designing PV installations in a "persona" point of view. Therefore, average profiles were created following the collected questionnaires' answers, and the PV installations were designed according to this profiling and the hourly electricity consumption data collected. Based on the results obtained, we launched a special sales campaign "all included" for PV systems to share with our clients and potential customers.

Figure 30: "All included" PV offer (flyer)

For the analysis of data, the process was, in first place, to collect the data (energy and questionnaire answers) through the SEAS and download the package to be deployed on an instance. Then, deploying all the packages and tools on a local instance. The SEAS has three main tools available: Jupyter, MLFlow and Apache Superset. For developing the scenario, A,

Jupyter and Superset were selected. Jupyter was used to process the files that store the energy data (in .json), to perform an energy consumption analysis per dataset (common to all the scenarios) and to design the PV installations. A Jupyter notebook was programmed, and the results (and raw data) were collected and automatically transferred to the SQL database from Apache Superset. Once the raw data and the results are stored in the SQL database, Superset was used to develop a general energy analysis (for all users) and to design Dashboards (one for the general energy analysis and another Scenario specific).

In the next figures there are shown some pieces of the code developed in Jupyter and the connection to the SQL database.

Figure 31: "Example of the structure of the .json file of a dataset

Figure 32: "Loading of datasets collected from DataVaults platforms.

```
#bucle while de calculo de autoconsumo
contador = 0
max_iteraciones = 1000
potencia = 0
paneles = 0
autoconsumo=np.zeros(len(datos_completos['inputActiveMagnitude']))
excedentes=np.zeros(len(datos_completos['inputActiveMagnitude']))
consumo_grid=np.zeros(len(datos_completos['inputActiveMagnitude']))
autoconsumo_pc = 1000
autoconsumo_optimo = 40

while (contador < max_iteraciones and autoconsumo_pc > autoconsumo_optimo and paneles < paneles_max)
    potencia += kWp_panel
    paneles = round(potencia/kWp_panel,0)
    datos_completos["generacion"] = datos_completos[orientacion].astype(float)*potencia
    contador += 1
    datos_completos["autoconsumo"] = datos_completos[['inputActiveMagnitude','generacion']].min(axis
    datos_completos["excedentes"] = (datos_completos["generacion"] - datos_completos["inputActiveMag
    autoconsumo_tot=datos_completos["autoconsumo"].sum()
    generacion_tot=datos_completos["generacion"].sum()
    #autoconsumo_pc=100*(generacion_tot/consumo_anual)
    autoconsumo_pc=100*(autoconsumo_tot/generacion_tot)

espacio_m2=paneles*m2_panel
coste_instalacion=round(potencia*coste_kWp*1000,2)

#bucle cálculo ahorros
datos_completos['ahorros']=np.zeros(len(datos_completos['inputActiveMagnitude']))
for i in range(len(datos_completos['periodo'])):
  if datos_completos['periodo'][i]==1:
    datos_completos["ahorros"][i]=datos_completos["autoconsumo"][i]*P1;
  else:
    if datos_completos['periodo'][i]==2:
        datos_completos["ahorros"][i]=datos_completos["autoconsumo"][i]*P2;
    else: #datos_completos['periodo'][i]==3:
```

Figure 33: “Piece of code of the calculation of the PV installation

```
•[11]: # Conexión a database PostgreSQL
conn = psycopg2.connect(database="mlflow-db", user="mlflow_user", password="mlflow_pwd", host="postgresc
cur = conn.cursor()
query = "DROP TABLE caso_1"
cur.execute(query)
query = "CREATE TABLE IF NOT EXISTS caso_1 (userID TEXT, potencia NUMERIC,paneles NUMERIC,espacio_m2 NUM
cur.execute(query)
totrec = 0
```

```
query = "INSERT INTO caso_1 (userID,potencia,paneles,espacio_m2,coste_instalacion, ahorro_total_anua
values = (userID, potencia,paneles,espacio_m2,coste_instalacion, ahorro_total_anual, tiempo_recup_in
cur.execute(query,values)

# Salvataggio delle modifiche nel database
conn.commit()
```

Figure 34: “Piece of code of the connection with SQL database and the transfer of results

RESULTS QUERY HISTORY

CREATE CHART DOWNLOAD TO CSV COPY TO CLIPBOARD Filter results

40 rows returned

userid	potencia	paneles	espacio_m2	coste_instalacion	ahorro_total_anual
ac4366d9-9785-41f2-bbb1-c2352ea81c00	1	2	4.4	1300	61.611630000000005
7f122797-af82-4671-b360-7eb5b61e8e4d	1.5	3	6.6000000000000005	1950	73.60035500000001
a3b3f528-9e5a-4c62-b5fb-0127da2d159d	2	4	8.8	2600	70.17888
b3a83135-ccba-4e0a-aeb3-796f4b2c08e9	1	2	4.4	1300	66.30389
0a822c46-17fa-4361-af48-d39f4e36c2ea	1	2	4.4	1300	69.18263
e4336d1b-b660-440c-a39e-174fe97a742a	2.5	5	11	3250	172.47673000000003

Figure 35: Results showed in SQL Lab from Apache Superset

General energy analysis and results

The next figure provides a screenshot of the dashboard designed in Apache Superset. The dashboard is presenting 7 graphs that analyse the total consumption, monthly energy consumption (globally and per user), and an hourly energy analysis per user. In addition, a calendar map has been designed in order to detect the days of the year with the highest energy consumption.

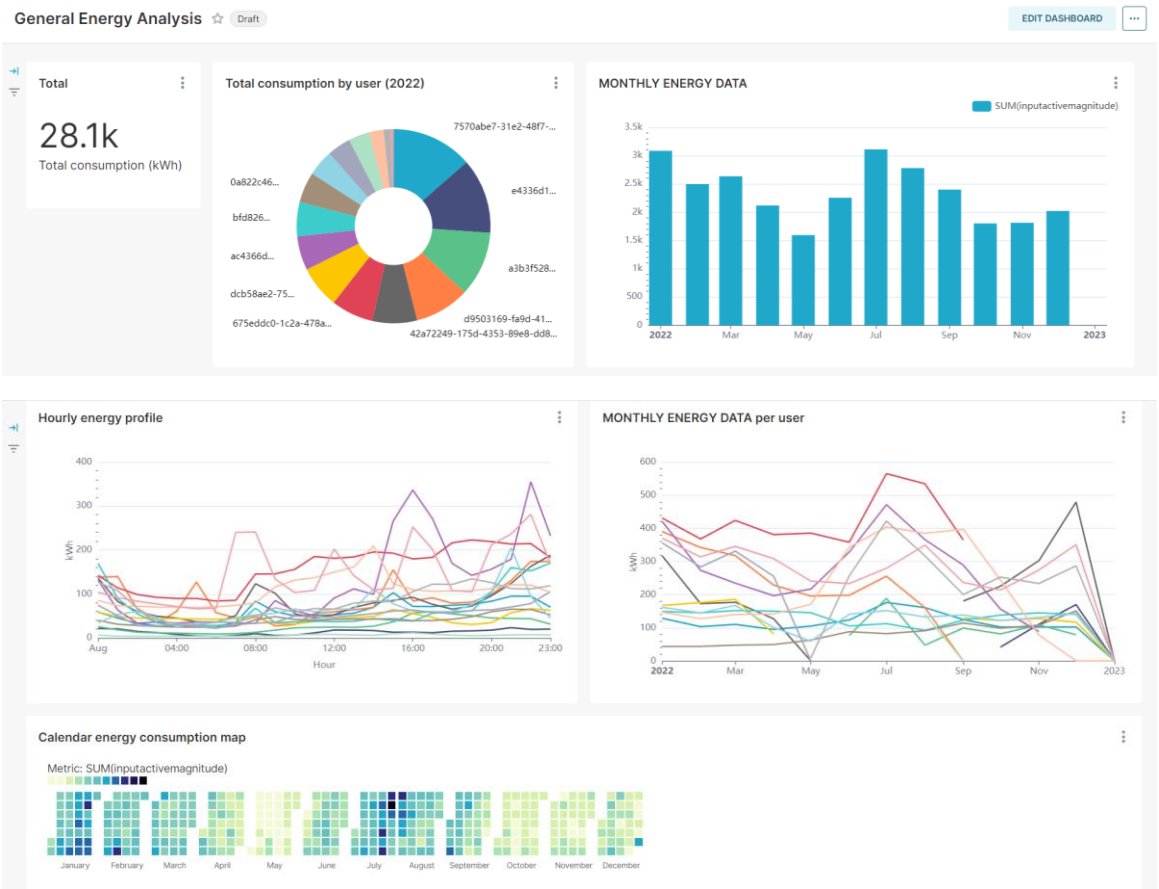

Figure 36: General energy analysis dashboard from Apache Superset

On the analysis of data, we have obtained that the total consumption is about 28.1 MWh, and that 37% of the total consumption is shared by only 3 users. The months with the highest consumption are January and July, probably because of the highest use of heating and cooling electrical appliances. This is also reflected in the calendar map, in which the days with the highest consumption are placed in July and January but they are also highlighted the 26th of February and the 1st of August. From the hourly energy consumption analysis, the main conclusion is that each user has very specific daily routines and habits that should be studied individually, but in most cases, there is a peak of consumption in the early morning hours (8h), Spanish lunch hour (15h) and dinner time (around 21-22h).

As can be seen, this type of analysis can help the Data Seeker to know more about the people that are sharing the data with them and study their energy consumption patterns, helping to understand their behaviour in order to design better services (for example, to design

electricity tariffs that take into account the hours in which the user normally consume more energy, offering better prices during this hours).

Analysis of data and results for scenario A

In first place, we analysed the answers received for the questionnaire and made an average profile of customer to design the PV installation. In the next graphs it is summarized the main results obtained from the questionnaire.

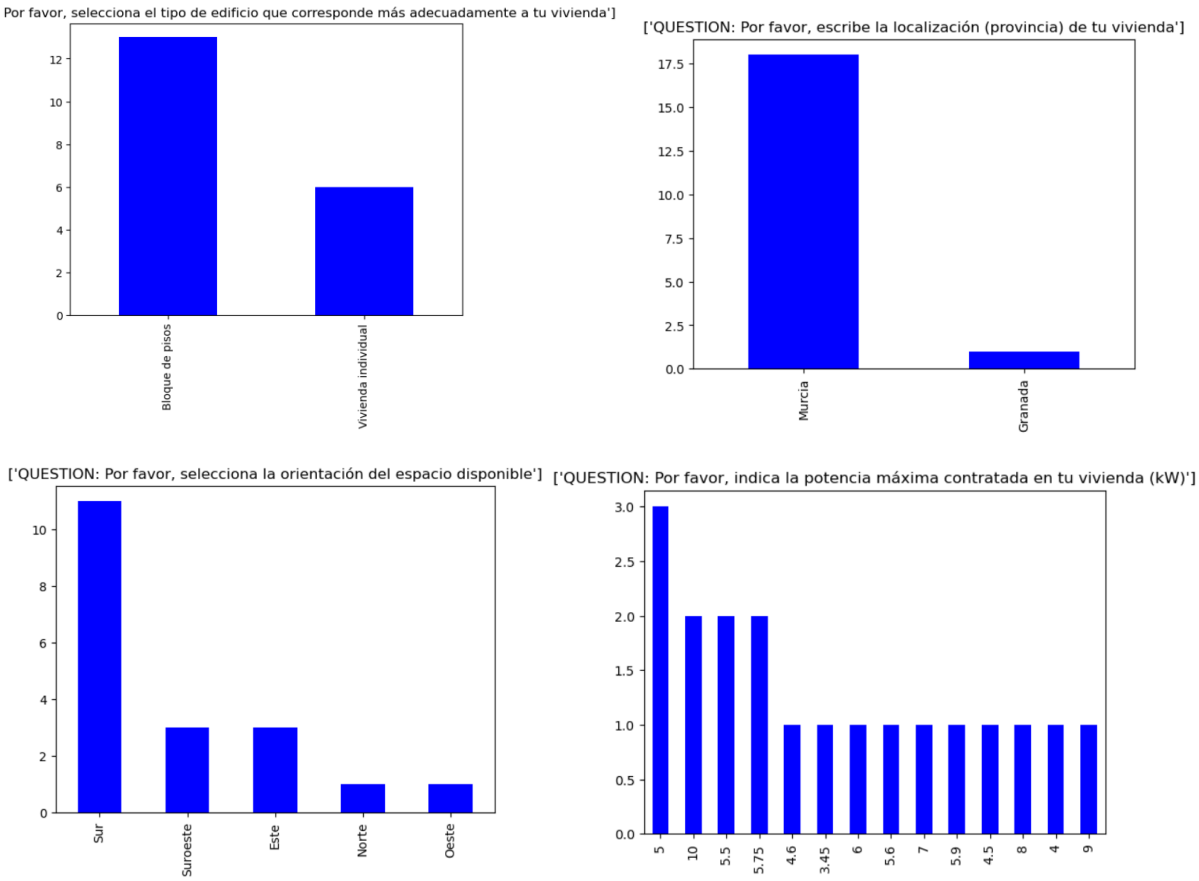

Figure 37: Questionnaire results for Scenario A

For the design of the PV installation, normally it is needed at least the yearly energy consumption to make estimations about the size of the plant. In this case, we have used the hourly energy consumption and we have calculated the optimized size of the plant taking into account a self-consumption rate of 40%. In the next figure it can be seen the dashboard that we have designed in Superset. The first two graphs represent the average energy consumption, PV production and self-consumption for one plant and the table shows the individual results per user.

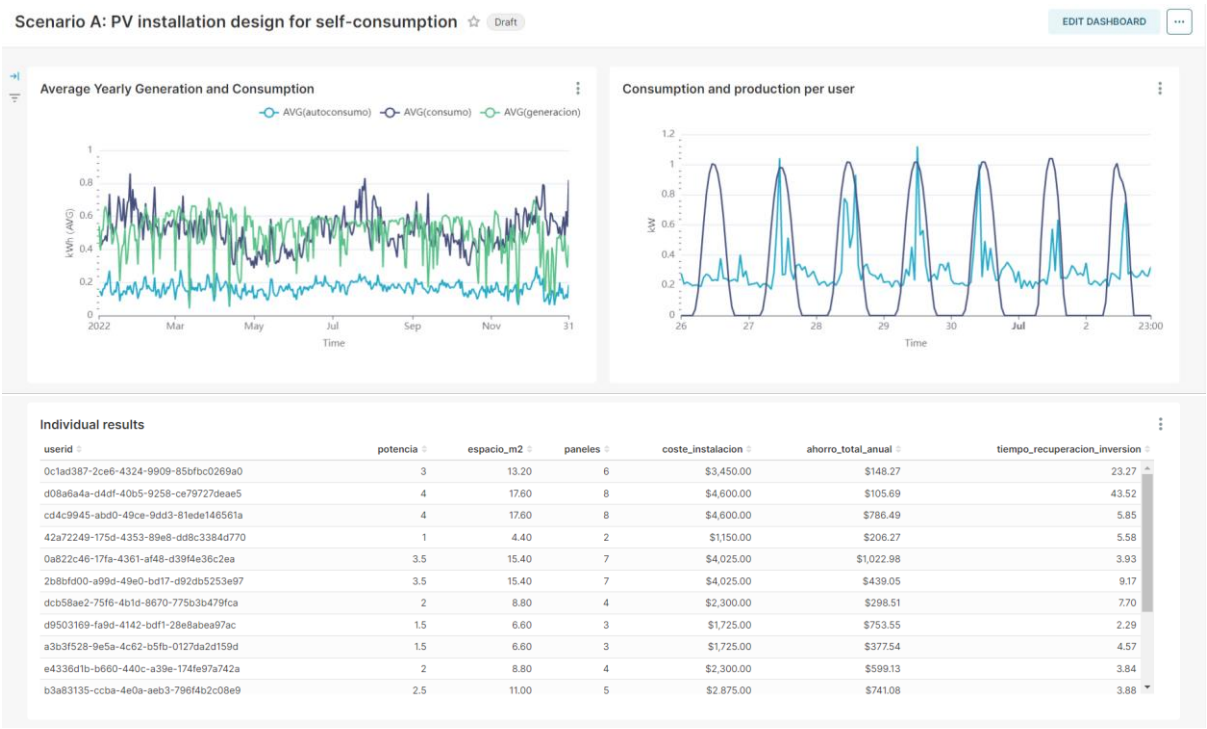

Figure 38: Scenario A dashboard from Apache Superset

It has been found that the profitability of a PV plant strongly depends on the consumption patterns of the client, as if there is not enough consumption during the solar hours, the profitability of the PV plants is very low. For this reason, the hourly energy analysis is very useful to improve the design of the PV installations and help to demonstrate to the client the adequacy of the installation. Although there is not enough questionnaire answers and datasets to make definitive conclusions, it can be deduced that the more interesting profile of a user that can install a profitable PV installation is a user that lives in an individual dwelling. However, small installations for people that live in buildings are also interesting and a market to explore. On the other hand, it is a key factor that the PV panels should be oriented to the South. Finally, the average profile of the customers that have shared their data are located in Murcia and they have a contracted power of around 6 kW.

3.4.2.3 *MIWENERGIA Scenario B - Improve profiling of the clients to enhance energy efficiency.*

3.4.2.3.1 Scenario Description

The *Scenario B* was defined with the main objectives of improving the knowledge of the energy sector and create awareness about energy efficiency among the society. In order to achieve this objective, a new energy certification/classification has been designed, based not on the building characteristics of the dwellings (like the existing certificates) but on the electricity consumption, and taking into account other personal data such as the number of inhabitants or the type of appliances they have.

3.4.2.3.2 Validity of Scenario

The scenario is still valid although some changes have been made in the approach to the final solution, being adapted to the final functionalities available in the platforms. As the questionnaires and the datasets cannot be related to which users have sent them, no tips on

energy efficiency can be sent to the specific user, but profiling and classifications based on energy consumption and dwellings characteristics have been done in order to create personalized services related to energy efficiency.

3.4.2.3.3 Objectives of MIWENERGIA Scenario B

Because some of the initial tasks are common to the different scenarios and have already been defined above. Only additional tasks will be mentioned here.

Objective	Status	Phase Achieved
Energy efficiency mock-ups definition	Achieved	Beta

Table 18: MiWenergia Scenario B objectives

3.4.2.3.4 MIWENERGIA Scenario B: Evolution

Scenario B was defined with the main objectives of improving the knowledge of the energy sector and creating awareness about energy efficiency among society. In order to achieve this objective, a new energy certification/classification has been designed, based not on the building characteristics of the dwellings (like the existing certificates) but on the electricity consumption, and taking into account other personal data such as the number of inhabitants or the type of appliances they have. To design this classification, the report[9] from the SPAHOUSEC *"Análisis del consumo energético del sector residencial en España" (Analysis of the energy consumption of the residential sector in Spain)* has been used as a reference. This report made an exhaustive analysis of the energy consumption in Spain, the energy sources and the characteristics of the buildings. Based on the data of the report, MIWenergia has designed a specific label per user profile and building characteristics.

In this case, the Data Seeker can be represented by any energy company that want to offer energy efficiency services. The certificate or classification of customers in relation with their energy consumption can help to design targeted marketing campaigns and personalized services. The Data Seeker can focus their efforts on the profiles that have worst ratings.

The evolution and development of the scenario follow the same schema than the previous one, so in this Section it is not going to be explained in detail again. The main tasks performed were:

- Planning of the scenario evolution and approach
- Design and development of the MIWenergia API.
- Definition of the tools and data sources needed.
- Definition of "extra" data needed and design of questionnaires.
- Definition of compensations methods.
- Testing of the platforms
- Recruiting, demo, trainings and support to participants
- Sharing of questionnaires with participants
- Buying and collection of data and questionnaire answers
- Analysis of data (general and scenario specific)

[9] https://www.idae.es/uploads/documentos/documentos_Informe_SPAHOUSEC_ACC_f68291a3.pdf

- Evaluation of results

In the next figures, there are screenshots of the questionnaires implemented in the platforms.

Figure 39: Questionnaire from the Cloud-Data Seeker perspective (Scenario B)

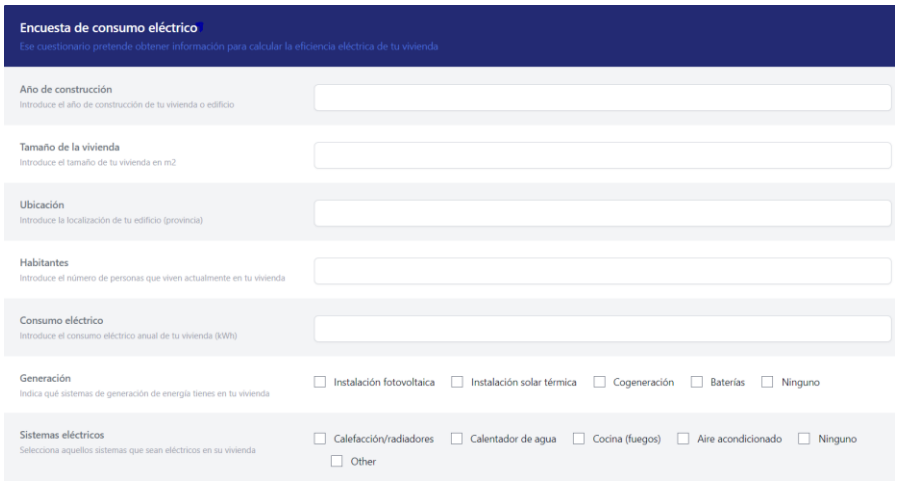

Figure 40: Questionnaire from the PersonalApp-Data Owner Perspective (Scenario B)

In the next paragraphs, the specific analysis of data and the results for Scenario B and the energy classification are explained in more detail.

Analysis of data and results for Scenario B

For performing the analysis of the data of Scenario B, the first step was to collect and process the answers to the questionnaire and the energy consumption data from the platform. In the next figures there are some screenshots from the code developed in Jupyter for processing the inputs from questionnaires and to upload the basic data from the SPAHOUSEC report.

```
zona_climatica=pd.read_csv('zona_climatica.csv', sep=';').set_index('PROVINCIA')
consumo_medio_appliances=pd.read_csv('consumo_medio_electrodomesticos.csv', sep=';').set_index('SERVICIO
categorias=pd.read_csv('categorias.csv', sep=';').set_index('CLASS')
superficie_media=pd.read_csv('superficie_media.csv', sep=';').set_index('Clima')
#etiqueta_m2=pd.read_csv('consumo_medio_m2.csv', sep=';').set_index('Etiqueta')
#etiqueta_total=pd.read_csv('consumo_medio_total.csv', sep=';').set_index('Etiqueta')

#Excel de datos de cuestionarios
respuestas = pd.read_excel("eficiencia_energetica.xlsx")
#codigos respuestas
    #Generación:0=FV; 1=SolarTermica, 2=Cogeneracion,3=Baterias,4=ninguno
    #sistemas electricos
    #calefaccion=0;WH=1;Cocina=2;HVAC=3; Ninguno=4
    #Electrodomesticos
    #frigo=0;congelador=1;lavadora=2;lavavajillas=3;Secadora=4; Horno=5; TV=6; PC=7;
```

Figure 41: Uploading the input data for Scenario B

```
cuestionario_dict=dict()
cuestionarios = list(pd.unique(respuestas["Cuestionario"]))
preguntas =  list(pd.unique(respuestas["Pregunta"]))
for cuestionario in cuestionarios:
    dicionario = dict()
    respuestas_cuestionario = respuestas[respuestas["Cuestionario"] == cuestionario]
    for pregunta in preguntas:
        respuestas_pregunta = respuestas_cuestionario[respuestas_cuestionario["Pregunta"] == pregunta]
        if len(respuestas_pregunta==1):
            dicionario[pregunta] = respuestas_pregunta["Answer"].iloc[0]
        else:
            dicionario[pregunta] = list(respuestas_pregunta["Answer"])
    cuestionario_dict[cuestionario]=dicionario
#print(cuestionario_dict)

cuestionario_list = list(cuestionario_dict.items())
```

Figure 42: Processing of the answers of questionnaires for Scenario B

After that, the program creates two classifications (energy labels) based on the characteristics of the user, its dwelling and the electrical system and appliances they have. The first classification is only based on the total consumption depending on the type of building, the electrical appliances they have and the climatic zone in which the building is located, and the second one also introduces the size of the building and the number of inhabitants as parameters to take into account. Once the classification is defined and based on the total consumption of the user, the two labels that correspond to the two classifications are obtained. The following figure shows a piece of code of the calculation of the two energy classifications and the label per user.

```
for userId, df in cuestionario_list:
    userIdd=userId
    print(userIdd)
    #Datos de cuestionarios
    usuario=userIdd.astype(np.str_).tolist()
    año_construccion=df['Año de construcción']
    area_vivienda=float(df['Tamaño de la vivienda'])
    provincia=df['Ubicación']
    clima=zona_climatica.loc[provincia]
    numero_habitantes=float(df['Habitantes']);

    consumoanual=float(df['Consumo eléctrico'])
    prueba=random.choice(consumos)
    arr=[consumoanual,prueba]
    consumo_anual=random.choice(arr)

    sistemas_generacion=[df['Generación']]
    sistemas_electricos=[df['Sistemas eléctricos']]
    appliances=[df['Electrodomésticos']]
    tipo_vivienda=df['Tipo de vivienda']#0=Pisos, 1=Unifamiliar, #2=Indeterminado/Total
```

```
    if tipo_vivienda==0:
        vvivienda='Pisos'
    elif tipo_vivienda==1:
        vvivienda='Unifamiliar'
    else:
        vvivienda='Total'

    climaa=clima['Zona Climatica']
    superficie_med=superficie_media[vvivienda]
    area_media=superficie_med[climaa]
    consumo_medio_m2=consumo_medio_actual/area_media
    consumo_m2=consumo_anual/area_vivienda
    consumo_habitante=consumo_anual/numero_habitantes
    print("Consumo por m2: {:.2f} kWh" .format(consumo_m2))
    categorias['Consumo']=categorias['less than'] * consumo_medio_actual
    categorias['Consumo_m2']=categorias['less than'] * consumo_medio_m2

    consumo_anual_arr=np.array(consumo_anual)
    for i in range(len(categorias)):
      if (categorias.Consumo[i] > consumo_anual_arr):
        etiqueta=categorias.index[i]
        break
    etiquetas.append(etiqueta)

    consumo_m2_arr=np.array(consumo_m2)
    for i in range(len(categorias)):
      if (categorias.Consumo_m2[i] > consumo_m2_arr):
        etiqueta_m2=categorias.index[i]
        break
    etiquetasm2.append(etiqueta_m2)
```

Figure 43: Calculation of energy classification and label per user for Scenario B

Once the classification is performed and the label per user is obtained, the results are sent to the Apache Superset environment (SQL database), in the same way that Scenario A, in order to create a Dashboard to summarize and graphically represent the results. In the next figure, there is a caption of the Dashboard designed for this Scenario.

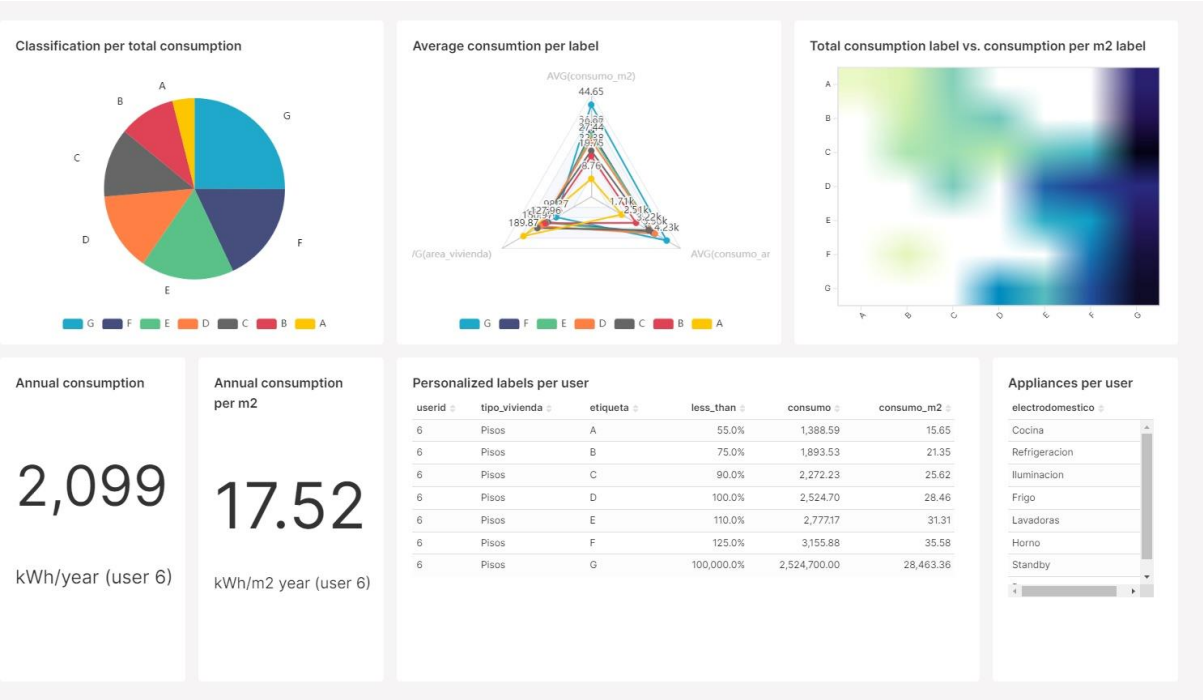

Figure 44: Scenario B dashboard from Apache Superset

As it can be seen most of the users have the worst energy efficiency labels related to the total energy consumption (letters from G to E). However, from the heat map, it can be deduced that many of the users that have the G class in relation to the total energy consumption have a C letter in relation to the energy consumption per square meter and inhabitant. This indicated that in many cases, although the first indicator of their efficiency said that their energy efficiency is low, it is necessary a more exhaustive analysis, taking into account more factors, to know if their house and their energy consumption behaviour are in fact efficient, and that is the reason for creating the second classification, and the specific classification per user and not a general one with average values. In the second row of the dashboard, there are represented the results for one specific user, in this case, user 6, in which it is shown the total yearly consumption, the consumption per m2, the personalized classification according to the characteristic of the user and the electrical appliances they have in their dwelling, according to the questionnaire answers. In this case, the labels that correspond to these results are C per total consumption and B per m2 consumption. This analysis can be done for each user and can be related to other characteristics such as the type of building or other personal characteristics, and in this way, create targeted campaigns to offer personalized energy efficiency services. This step will be done in the future when more people are involved in the platform.

3.4.2.4 *MIWENERGIA Scenario C – Energy consumption patterns with personal preferences*

3.4.2.4.1 Scenario C: Scenario Description

The *Scenario 3: Energy consumption patterns with personal preferences* was designed with the main objective being to design personalized services or offers related to the electricity supply but considering also other personal information such as hobbies, preferences or tastes

of the clients and potential clients. These personal data are not easy to obtain for MIWenergia due to the status and characteristics of the company, so the marketing department lacks this information in the design of the marketing campaigns. DataVaults platform can contribute to fill in this gap. At the same time, Data Owners will only share the information they want and in the way they prefer (anonymized, encrypted, within a persona, etc.), receiving compensations in exchange for it, so the way to obtain the data is in concordance with the ethics of the company. In addition, a forecasting model for hourly energy consumption data has been designed in this scenario with two objectives: to evaluate energy consumption patterns and demand prediction and to compare the results of the model with the ones that are currently subcontracted.

3.4.2.4.2 Validity of Scenario

The scenario remained still valid although some changes have been made in the approach to the final solution, being adapted to the final functionalities available in the platforms. Energy consumption prediction models have been implemented and profiling of personal preferences from users has been done.

3.4.2.4.3 Objectives of Demonstrator's Scenario

Objective	Status	Phase Achieved
Hobbies and interests mock-ups definition	Achieved	Final

Table 19: MiWenergia Scenario C objectives

3.4.2.4.4 MIWENERGIA Scenario C: Evolution

The evolution and tasks performed in this third scenario follows the same structure than the two previous ones, so they are not further explained here. In the next figures, there are screenshots from the questionnaire implemented in the platforms.

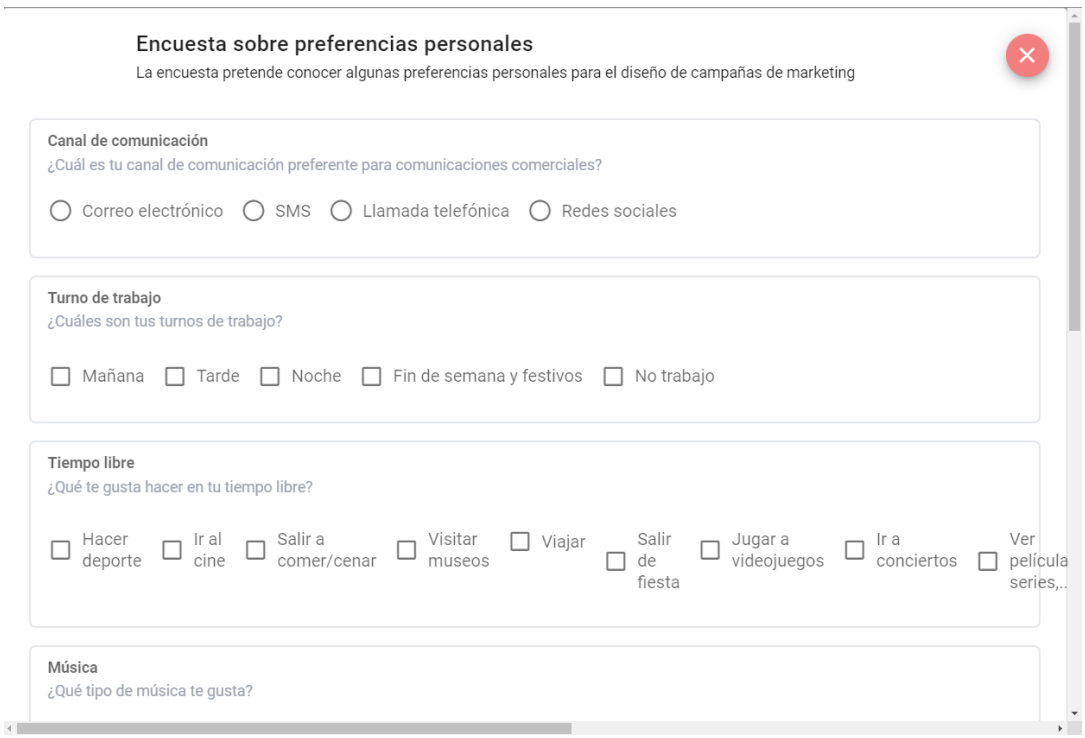

Figure 45: Questionnaire from the Cloud-Data Seeker perspective (Scenario C)

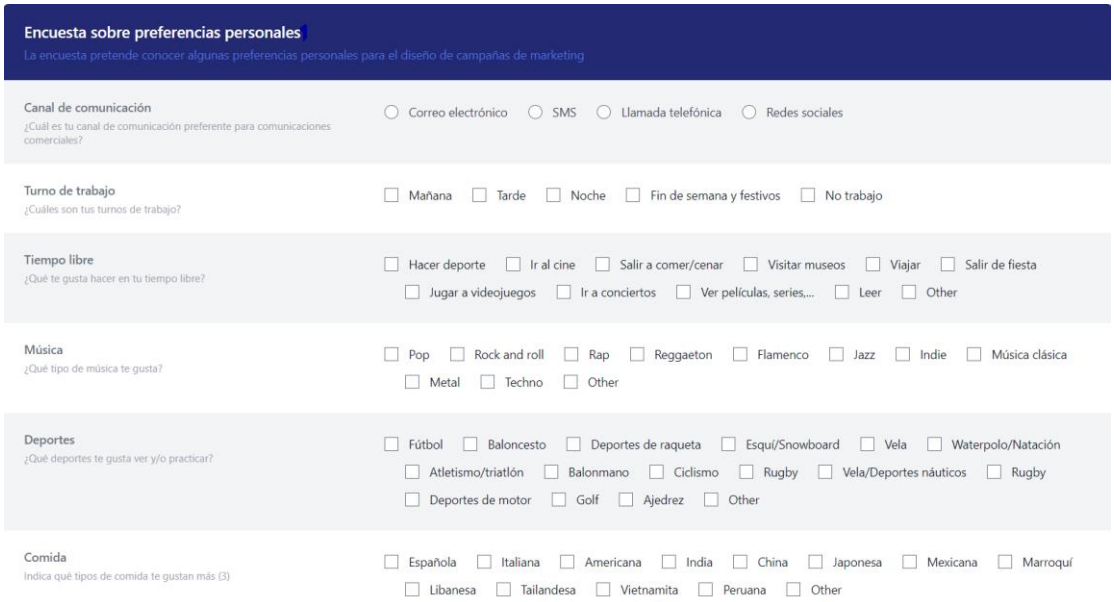

Figure 46: Questionnaire from the PersonalApp-Data Owner perspective (Scenario C)

In the next paragraphs it is described the prediction model developed and the results of the profiling of customers based on the questionnaires answers.

Analysis of data and results for Scenario C

Regarding the profiling of the users, the first step was to design a questionnaire of 8 basic questions to obtain information about the customers. The results of the questionnaire implemented in the cloud are shown in the next graphs. Most of the questions are targeted to design marketing campaigns, asking about preferences and hobbies, and some questions about daily habits to help in the design of electricity tariffs of the company. The results show that 21 of 23 people that answer prefer to receive commercial communications via email, so the next campaigns of the company will be focused on this channel. On the other hand, 65% of the users prefer to have direct discount codes, 22% prefer to receive gifts and the rest prefer exchangeable vouchers for activities. In order to design the gifts or vouchers, some questions have been asked that reflect that the activities that they normally do in their free time are going out for dinner/lunch, travelling and working out/playing sports. For dinner and lunch, they are keener on Spanish food, but there is also interest in Italian and Japanese. In the case of the sport, the most liked is football, followed by racquet sports, basket and cycling. Regarding entertainment, the most liked music genre is pop, and mystery is preferred for series and films. Finally, the participants mainly work in the morning shift although there is a significant amount that also work in the afternoon. This information can be used to design specific tariffs with lower prices when people are at home and higher prices when are out working. In addition, there can be defined also a tariff for people that work remotely from home, although in this case, this specific question has not been asked.

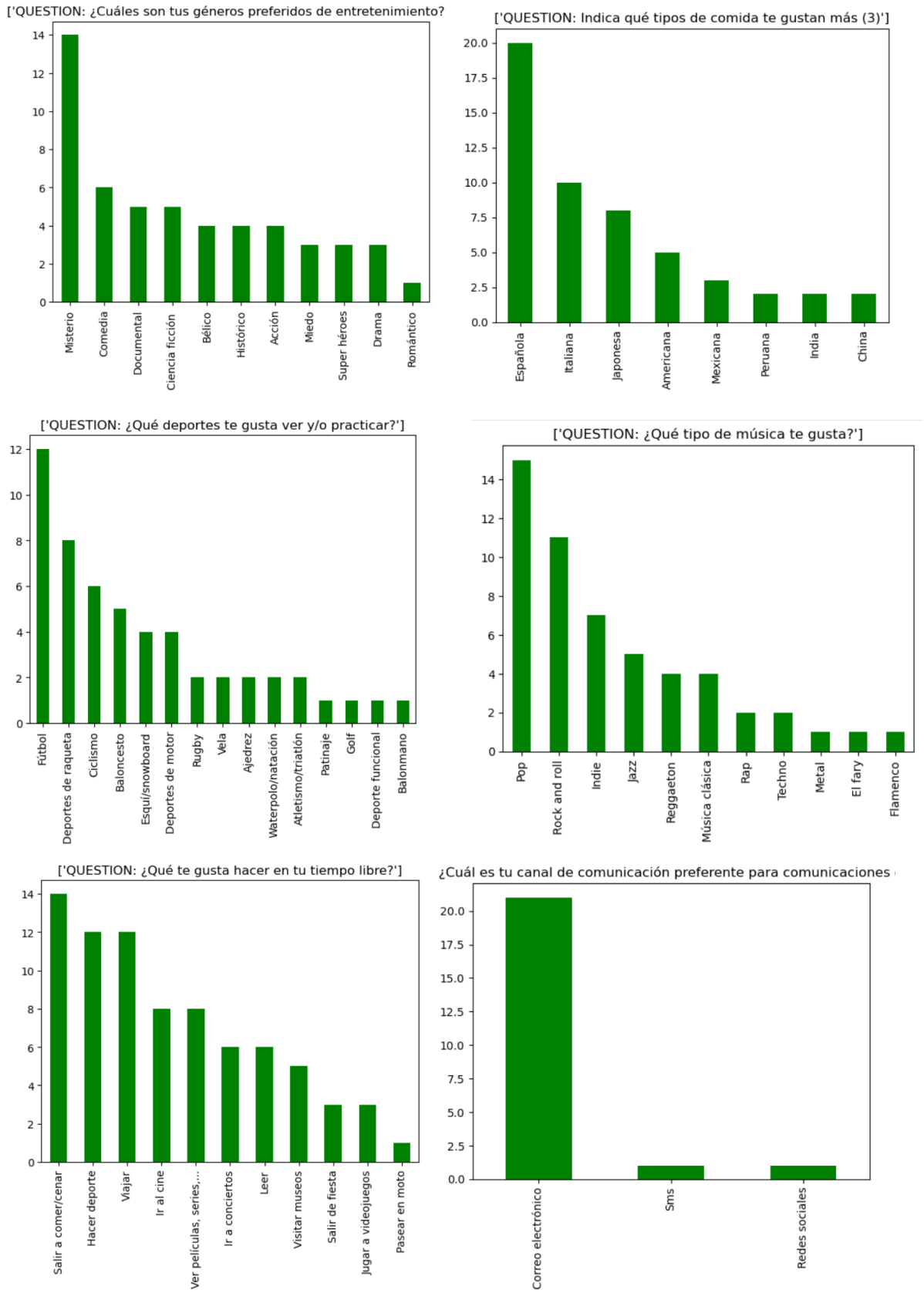

Figure 47: Questionnaire results from Scenario C (I)

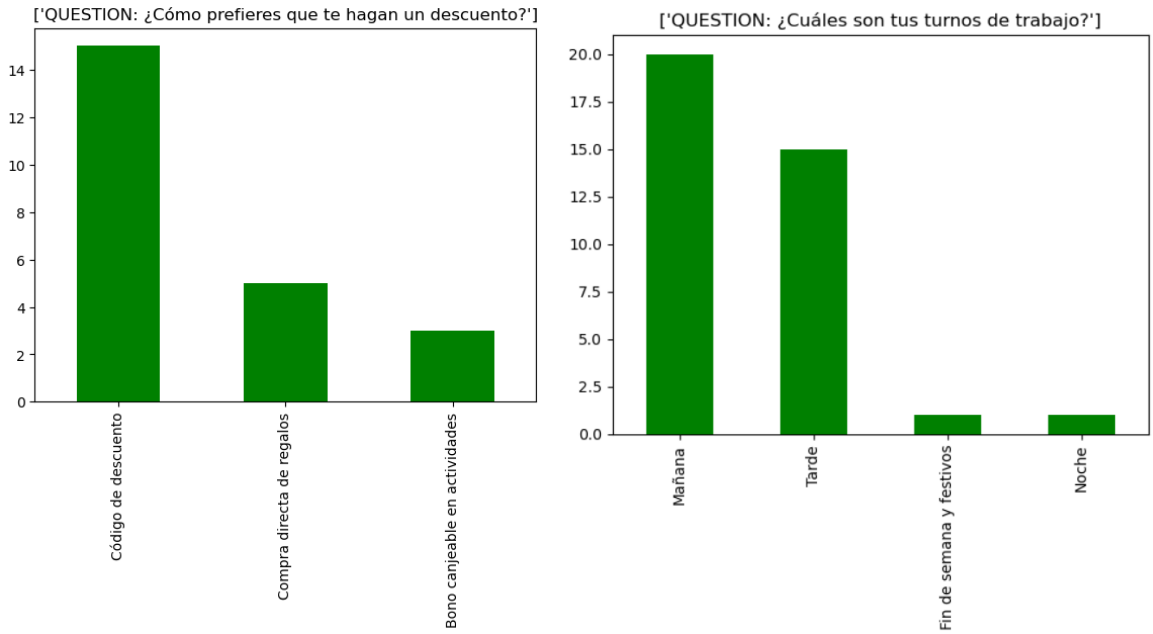

Figure 48: Questionnaire results from Scenario C (II)

On the other hand, for the development of a forecasting model for the total hourly energy consumption of the participants, the Jupyter tool has been used. Currently, MIWenergia has subcontracted the forecasting service to a third party, so the objective of this scenario was to test models that can replace this service.

We have tested two types of multi-step forecasting models from the library skforecast of Python: a recursive multi-step forecasting model and a direct multi-step forecasting model. The model finally selected was the direct multi-step model. This type of model is computationally more expensive than the recursive since it requires training several models, but in this case, it shows better results. The model also includes exogenous variables, such as the day of the week and the holidays.

The main steps to develop the model were:

1. Analysis and processing of the input data
2. Preliminary and graphical analysis of data
3. Definition of the exogenous variables
4. Training and tuning of the forecaster.
5. Backtesting predictions
6. Evaluation of the backtest error.

In the next paragraphs, there are explained the results and the conclusions that have been obtained and there are also shown some screenshots of the code developed in Jupyter.

On the preliminary analysis and processing of data, we collect all the datasets that MIWenergia have bought from the DataVaults platform, and we have aggregated them. We only have data from 2022 and around 20 users (although some users have more than one supply point).

```
for userId, df in df_list:

    userID=userId
    df = df.reset_index()
    del(df['isDefinitive'])
    del(df['quadrant1ReactiveMagnitude'])
    del(df['quadrant2ReactiveMagnitude'])
    del(df['quadrant3ReactiveMagnitude'])
    del(df['quadrant4ReactiveMagnitude'])
    del(df['uspc'])
    del(df['outputActiveMagnitude'])
    del(df['index'])

    kW=[]
    for numero in df.inputActiveMagnitude:
      kW.append(float(numero))

    df['Demand']=kW
    del(df['inputActiveMagnitude'])

    if not isinstance(df['isSummerTime'].iloc[0], str):
        df['isSummerTime'] = df['isSummerTime'].apply(str)

    df['isSummerTime'] = df['isSummerTime'].str.lower().str.contains('true')

    for aa in range(len(df)):
        if df.isSummerTime[aa]==True:
            df['timeStamp'][aa] += ' +0200'
            #print(df['timeStamp'][aa])
        else:
            df['timeStamp'][aa] += ' +0100'
            #print(df['timeStamp'][aa])
```

```
    del(df['isSummerTime'])
    df['Time'] = pd.to_datetime(df['timeStamp'], format='%Y-%m-%dT%XZ %z',utc=True)
    df['Time'] = df['Time'].apply(lambda x: x.strftime('%Y-%m-%d %H:%M:%S'))
    df['Time'] = pd.to_datetime(df['Time'],format='%Y-%m-%d %H:%M:%S')
    df = df.set_index('Time')
    df = df.sort_index()

    df=df.groupby(df.index).sum()
    df = df.asfreq('1H')
```

```
    # Verificar que un índice temporal está completo y completar huecos
    # ==============================================================================
    (df.index == pd.date_range(start=df.index.min(),end=df.index.max(),freq=df.index.freq)).all()

    df.asfreq(freq='1H', fill_value='bfill')

    consumos.append(df)
    #consumos=consumos.groupby(consumos.index).sum()

consumos_concat = pd.concat(consumos)

# Agrupar por fecha y sumar los valores de consumo
datos = consumos_concat.groupby(consumos_concat.index).sum()
```

Figure 49: Processing of input data from Scenario C (I)

The total number of data points was 8760, that were divided into 3 sub-parts for developing the forecasting model: train(n=5807), validation (n=2208) and test (n=744) datasets.

```
[15]: # Separación datos train-val-test
      # ==============================================================================
      datos = datos.loc['2022-01-01 00:00:00': '2022-12-31 23:59:00']
      fin_train = '2022-08-30 23:59:00'
      fin_validacion = '2022-11-30 23:59:00'
      datos_train = datos.loc[: fin_train, :]
      datos_val   = datos.loc[fin_train:fin_validacion, :]
      datos_test  = datos.loc[fin_validacion:, :]

      print(f"Fechas train      : {datos_train.index.min()} --- {datos_train.index.max()}  (n={len(datos_train)})")
      print(f"Fechas validacion : {datos_val.index.min()} --- {datos_val.index.max()}  (n={len(datos_val)})")
      print(f"Fechas test       : {datos_test.index.min()} --- {datos_test.index.max()}  (n={len(datos_test)})")
```

```
Fechas train      : 2022-01-01 01:00:00 --- 2022-08-30 23:00:00  (n=5807)
Fechas validacion : 2022-08-31 00:00:00 --- 2022-11-30 23:00:00  (n=2208)
Fechas test       : 2022-12-01 00:00:00 --- 2022-12-31 23:00:00  (n=744)
```

Figure 50: Division of the original dataset on train, validation and test sub-datasets

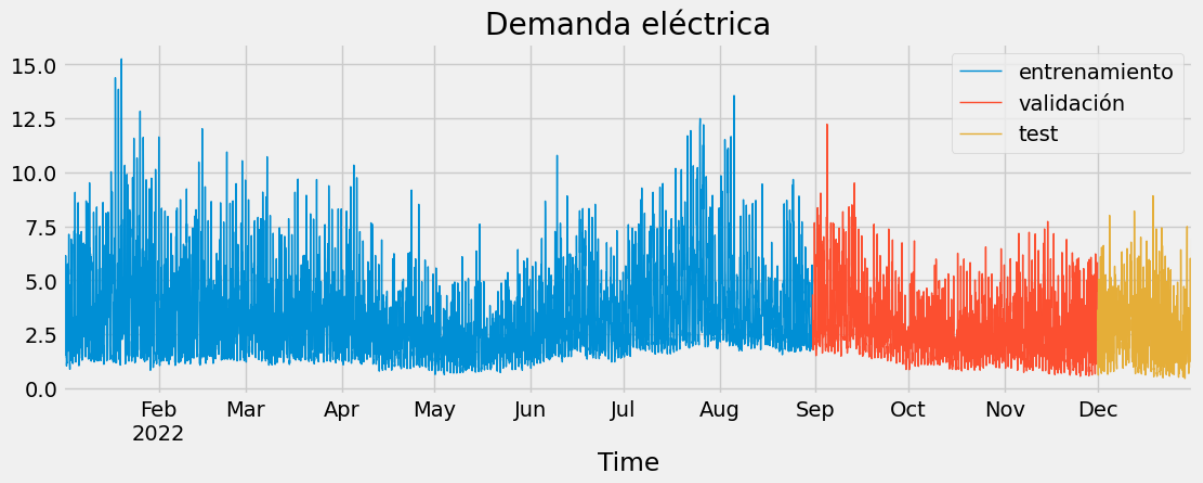

Figure 51: Total consumption dataset divided on train(blue), validation(red) and test (yellow) data.

In addition, an analysis of the seasonality of the data (monthly, weekly and hourly) has been performed. The monthly analysis made only sense if more than 1 year of data is available, but we have programmed it for future analysis or other longer datasets. Results show that the consumption increases significantly during the summer and at night (from 21-23h). The weekly analysis does not show any clear pattern, although the weekends' consumption is slightly lower. Graphical results are shown in the next figure.

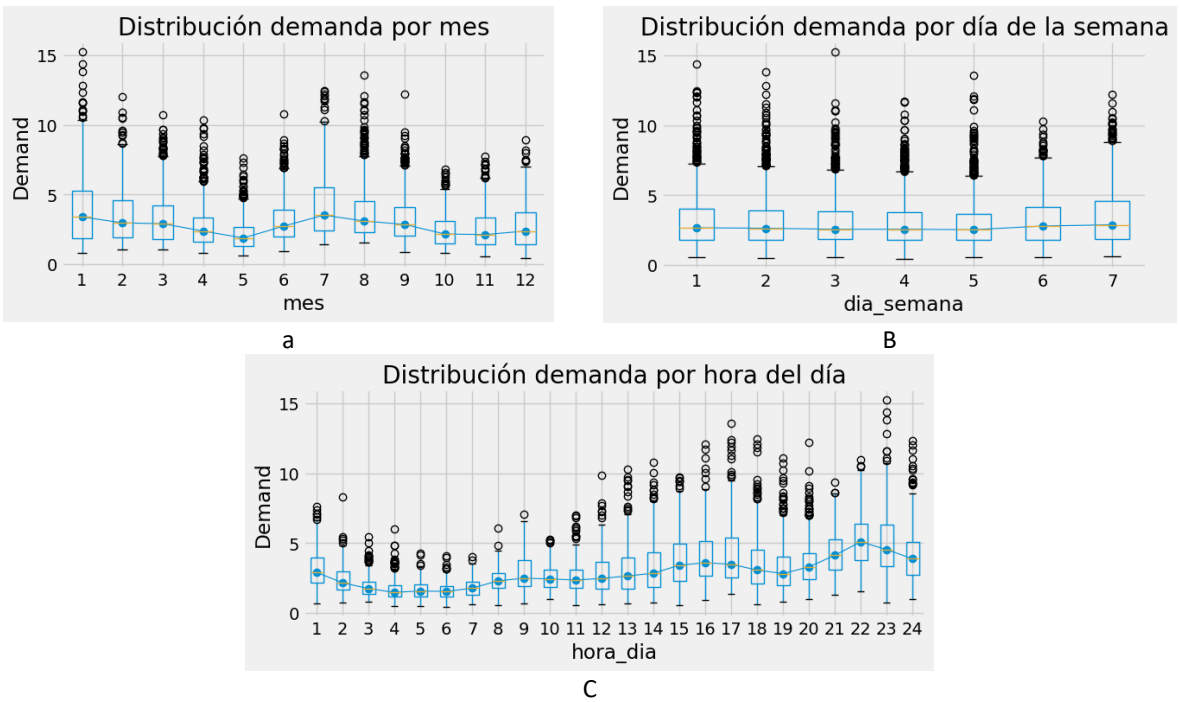

Figure 52: Seasonality analysis a. Monthly b. Weekly c. Hourly

Finally, an autocorrelation analysis has been performed, obtaining that there is a strong correlation with the lags 1,2,3,22,23,24,25,47,48,49. This makes sense as the consumption of a concrete hour of a day is strongly related to the consumption of the last hours of the same day and the same hours of the previous days.

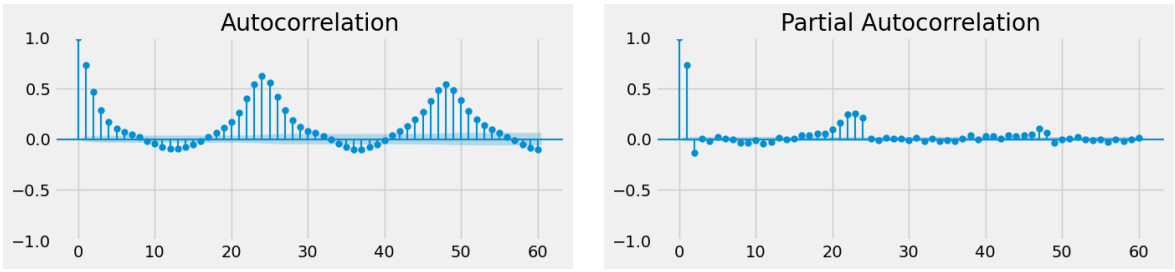

Figure 53: Autocorrelation analysis a. Direct b. Partial

Based on this preliminary analysis, an initial recursive forecaster has been configurated and tuned, and an evaluation of the lags' importance has been performed.

```
[51]: # Importancia predictores
      # ==============================================================================
      forecaster.get_feature_importance().sort_values('importance',ascending=False)
```

[51]:

	feature	importance
0	lag_1	1572
23	lag_24	805
47	lag_48	553
22	lag_23	522
21	lag_22	502

Figure 54: Evaluation of lags importance

Moreover, the definition of the backtest for the hourly predictions has been coded.

```
[47]: def backtest_predict_next_24h(forecaster, y, hour_init_prediction, exog=None,
                                   verbose=False):

          '''
          Backtest ForecasterAutoreg object when predicting 24 hours of day D+1
          statring at specific hour of day D.

          Parameters
          ----------
          forecaster : ForecasterAutoreg
              ForecasterAutoreg object already trained.

          y : pd.Series with datetime index sorted
              Test time series values.

          exog : pd.Series or pd.Dataframe with datetime index sorted
              Test values of exogen variable.

          hour_init_prediction: int
              Hour of day D to start predicciones of day D+1.


          Returns
          -------
          predicciones: pd.Series
              Value of predicciones.

          '''

          y = y.sort_index()
          if exog is not None:
              exog = exog.sort_index()
```

Figure 55: Backtest predictions definition (I)

```
    dummy_steps = 24 - (hour_init_prediction + 1)
    steps = dummy_steps + 24

    # First position of `hour_init_prediction` in the series where there is enough
    # previous window to calculate lags.
    for datetime in y.index[y.index.hour == hour_init_prediction]:
        if len(y[:datetime]) >= len(forecaster.last_window):
            datetime_init_backtest = datetime
            print(f"Backtesting starts at day: {datetime_init_backtest}")
            break

    days_backtest = np.unique(y[datetime_init_backtest:].index.date)
    days_backtest = pd.to_datetime(days_backtest)
    days_backtest = days_backtest[1:]
    print(f"Days predicted in the backtesting: {days_backtest.strftime('%Y-%m-%d').values}")
    print('')
    backtest_predicciones = []

    for i, day in enumerate(days_backtest):
        # Start and end of the last window used to create the lags
        end_window = (day - pd.Timedelta(1, unit='day')).replace(hour=hour_init_prediction)
        start_window = end_window - pd.Timedelta(forecaster.max_lag, unit='hour')
        last_window = y.loc[start_window:end_window]

        if exog is None:
            if verbose:
                print(f"Forecasting day {day.strftime('%Y-%m-%d')}")
                print(f"Using window from {start_window} to {end_window}")

            pred = forecaster.predict(steps=steps, last_window=last_window)
        else:
            start_exog_window = end_window + pd.Timedelta(1, unit='hour')
            end_exog_window   = end_window + pd.Timedelta(steps, unit='hour')
            exog_window = exog.loc[start_exog_window:end_exog_window]
            exog_window = exog_window

            if verbose:
                print(f"Forecasting day {day.strftime('%Y-%m-%d')}")
                print(f"    Using window from {start_window} to {end_window}")
                print(f"    Using exogen variable from {start_exog_window} to {end_exog_window}")

            pred = forecaster.predict(steps=steps, last_window=last_window, exog=exog_window)

        # Only store predicciones of day D+1
        pred = pred[dummy_steps:]
        backtest_predicciones.append(pred)

    backtest_predicciones = np.concatenate(backtest_predicciones)
    # Add datetime index
    backtest_predicciones = pd.Series(
                                data  = backtest_predicciones,
                                index = pd.date_range(
                                            start = days_backtest[0],
                                            end   = days_backtest[-1].replace(hour=23),
                                            freq  = 'h'
                                        )
                            )

    return backtest_predicciones
```

Figure 56: Backtest predictions definition (II)

The next step was the creation and tunning of the direct multi-step autoregressive forecaster.

```
[59]: # Grid search de hiperparámetros
      # ======================================================================
      forecaster = ForecasterAutoregDirect(
                      regressor = LGBMRegressor(max_depth=4),
                      steps     = 36,
                      lags      = 24 # Este valor será remplazado en el grid search
                  )

      # Lags utilizados como predictores
      lags_grid = [[1, 2, 23, 24], [1, 2, 3, 23, 24, 25, 47, 48, 49],[1,2,22,23,24,47,48]]

      # Hiperparámetros del regresor
      param_grid = {'n_estimators': [100, 500],
                    'max_depth': [4, 6]}

      resultados_grid = grid_search_forecaster(
                              forecaster         = forecaster,
                              y                  = datos.loc[:fin_validacion, 'Demand'],
                              exog               = datos.loc[:fin_validacion, exog],
                              param_grid         = param_grid,
                              lags_grid          = lags_grid,
                              steps              = 36,
                              metric             = 'mean_absolute_error',
                              refit              = False,
                              initial_train_size = len(datos.loc[:fin_train]),
                              return_best        = True,
                              verbose            = False
                          )
```

Figure 57: Creation and tunning of the forecaster

```
[60]: # Backtest
      # ======================================================================
      predicciones = backtest_predict_next_24h(
                          forecaster = forecaster,
                          y          = datos.loc[fin_validacion:, 'Demand'],
                          exog       = datos.loc[fin_validacion:, exog],
                          hour_init_prediction = 11,
                          verbose    = False
                      )
```

Figure 58: Backtesting

Once the forecaster is tuned, the backtest predictions have been performed and the results for the test data are presented in the following figure.

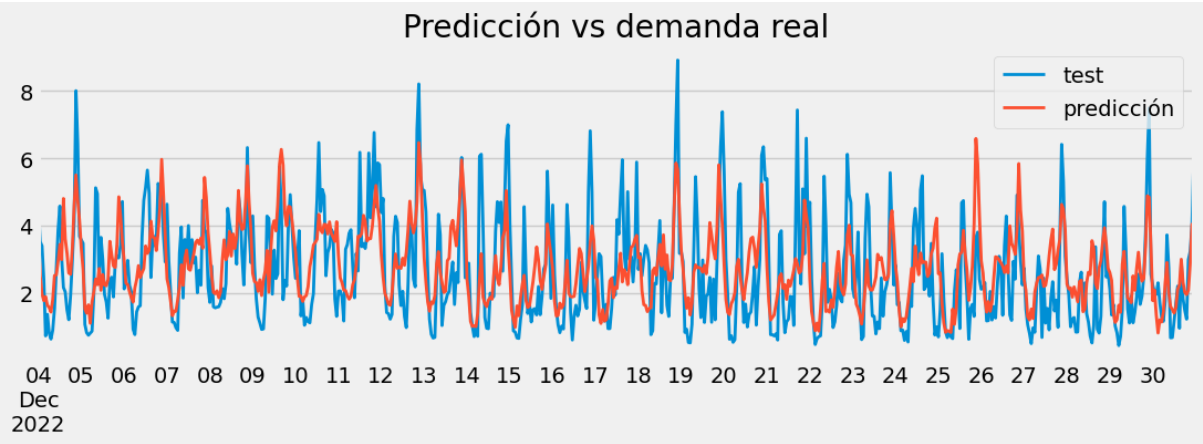

Figure 59: Results of the prediction model (red) vs real consumption (blue)

Finally, the mean absolute error and the root mean squared error has been calculated for the predictions obtained from the forecaster.

```
[63]: # Error backtest
MAE = mean_absolute_error(y_true = datos.loc[predicciones.index, 'Demand'], y_pred = predicciones)
print(f"MAE of backtest: {MAE}")

RMSE = math.sqrt(mean_squared_error(y_true = datos.loc[predicciones.index, 'Demand'],y_pred = predicciones))
print(f"RMSE of backtest: {RMSE}")

MAE of backtest: 1.004641041442561
RMSE of backtest: 1.2518482820083208
```

Figure 60: Error evaluation of the backtest

It has been obtained a MAE of 1.0 MW and a RMSE of 1.25. This error is high compared with other similar models, but the main reason is the low number of datasets from customers that have been obtained through the platform. The prediction of individual consumption is highly dependent on customer habits, but when a great number of customers are aggregated, the patterns are more stable and predictable. For this reason, a strong conclusion cannot be made about the adequacy of the model, which should be tested with a higher amount of datasets. Moreover, other exogenous variables should be studied and tested, such as the meteorological predictions and the number of clients.

MIWenergia will use this initial model to start the development of its own model for the portfolio of its clients, in order to create its own prediction model and implement it when improving the predictions from the subcontracted services.

3.4.3 High Level Impact Recorded in the Demonstrator

3.4.3.1.1 Benefits for the Data Seekers recorded from the Scenario.

The main Data Seeker of the three scenarios was MIWenergia. MIWenergia can benefit from the use of the platform via the collection of personal and energy data that can help the company to design personalized services, special targeted offers, and for the profiling of clients and potential clients.

In addition, two more potential Data Seekers have been reached. The first one is a public university, interested in collecting data for research purposes. They have registered in the platform and made some demonstration activities, creating and sharing some questionnaires and buying some datasets. The second Data Seeker reached is a private company, a PV installer. They are not interested directly in collecting data from the platform but in the results from the analysis of energy data that MIWenergia have done in Scenario 1 (e.g., could be classified as a tier-2 economic operator) of the demonstrator. In this way, MIWenergia has sent them the results of this scenario and a discussion on a potential collaboration between the two entities have started.

In general, any energy-related company can benefit from the demonstration performed in these case studies, and specially the PV installations companies can take advantage of the Scenario A. In addition, energy data can be very useful for other institutions such as research bodies (like the university that tested the platform as a Data Seeker) and public administrations (for example, for energy planning in smart cities, for facilitating renewable integration and development, for the design of energy communities, etc).

3.4.3.1.2 Benefits for the Individuals recorded from the Scenario.

The DataVaults project and the Demonstrator #4 have helped the participants to be much more aware of the importance of the privacy on data, as well as the value that its personal data have. In this way, the wallet and compensation methods of the platform helps users to understand this issue, and to set a price for their data. Although there has not been a lot of time to test the platform, especially the wallet and compensation methods, now the participants know that the personal information they can provide has a real economic value. Moreover, DataVaults project has contributed to a better understanding of the privacy terms and policies which are not well-known by the common citizen.

3.4.3.1.3 Benefits for the Organisation recorded from the Scenario.

As an organization, the three scenarios have helped MIWenergia to explore new tools that can be useful in the operation of the company, and to design new methodologies to create marketing campaigns and personalized services. In addition, the company has gained experience and knowledge on the data privacy and data analysis. Other benefits that can be highlighted from the development of the project are:

- Establishment of collaborations with new entities and improvement of the relationship with existing partners.
- Development of an API to extract energy data from clients and non-clients in an easy way. This API it is being evaluated to be integrated in the operation of the company after the end of the project.
- Creation of closer and loyal relationships with the participants of the project.

The benefits associated to the KPIs defined on the DoA are presented and discussed in the next table.

KPI	Initial value	Goal	Data collection method	Current Value	% of achievement
More effective management of customers.	60%	80%	Annual surveys	87,3%	136,5%
Increase in revenue through offering personalised services.	0€	30.000€	Annual economic report	Not possible to evaluate	N/A
Increase in revenue through sales agreements	0€	20.000€	Annual economic report	15.495€	77,5%
Increase in the number of partners.	20	30	Annual economic report	35	150%
Increase in client's satisfaction and trust.	70%	95%	Annual surveys	92,6%	90,4%
Increase in partners' satisfaction and trust.	80%	95%	Annual surveys	93,1%	86,7%

Table 20: KPIs for MIWENERGIA operating the platform.

From the table we can see that in the case of the effective management of customers and the increase in number of partners, the initial goal has been surpassed. On the other hand, in the case of the clients and partners satisfaction and trust, the objective has not been fully

achieved, although the current value is very close to the target. Finally, for the revenues, the DataVaults project have helped the company to sell two PV installations to two of the customers that have been participating for a total value of around 15.500€. However, there has not been enough time to implement personalized services designed on the basis of the project's results, so it has not been possible to evaluate this KPI.

3.4.3.1.4 Demonstrator KPIs

In this section, we present the KPIs relevant to the demonstrator as identified in deliverables D6.1 and D6.2, as well as other KPIs recorded during the demonstration phase.

It is noted, that not all these KPIs could be measured during the demonstrator phase of the project, as some of them concern post-project targets, as part of the exploitation activities of the demonstrator. Those are marked as such "Long-Term Impact" KPIs and are discussed in deliverable D7.4

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Status at M40	Comments
More effective management of customers.	We consider that the KPI should be based on incremental amounts referred to the incremental expected. Currently at 60%	Increase to 80% We propose and index α4, EMC = Effective Management of customers at the time, in % The index 4 refers to the number of demonstrators.	Currently from surveys.	Achieved	
Increase in revenue through offering personalised services.	Currently none.	0€ to 30,000€ We propose and index β4, RPS = Revenues through offering personalized services	in € from the economic reporting	Not implemented	No new personalized services related to the DV project have been launched as the demonstration phase has not been long enough.
Increase in revenue through sales agreements	Currently none.	0 to 20.000€ We propose and index π4, RSA = Revenues through sales agreements.	in € from the economic reporting	77% achieved	

Increase in the number of partners.	20 currently.	Increase to 30. We propose and index δ4, NOP = Number of partners	Units from the economic reporting	Achieved	15 new partners
Increase in client's satisfaction and trust.	Clients' satisfaction and trust is currently at 70%	To increase to 90% We propose and index Ω4, CST= Client's satisfaction and trust, in %.	From surveys	Almost achieved	
Increase in partners' satisfaction and trust.	Currently at 80%	To increase to 95% We propose and index 4, PST= Partners satisfaction and trust, in %.	From surveys	Almost achieved.	

Table 21: Demonstrator #4 KPIs

3.4.4 Demonstrator's Activities Timeline

The following table presents the main activities performed within the demonstrator.

Demonstrator 4 MIWENERGIA	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40
Scenario A – PV installation design for self-consumption																						
Alpha Phase																						
Testing MIWenergia API																						
API adaptation																						
Dwelling mock-up definition																						
Data verification																						
Participant Recruitment																						
Final phase																						
API, data reliability																						
Real users																						
Energy basic analyses																						
Data extraction																						
Implementation virtual wallet																						
Implementation of compensation methods																						
PV calculation method																						
Scenario B – Improve profiling of the clients to enhance energy efficiency																						
Alpha Phase																						
Energy efficiency mock-ups definition																						
Final phase																						
Clustering basic analyses																						
Data extraction																						
Energy efficiency factors																						
Scenario C – Energy consumption patterns with personal preferences																						
Alpha Phase																						
Hobbies and interests mock-ups definition																						
Final phase																						
Profiles and basic analyses																						
Data extraction																						
Def. Commercial campaigns																						
Playground external data test																						

Table 22: Execution Timeline for Demonstrator #4

3.5 DEMONSTRATOR #5 - PERSONAL DATA FOR MUNICIPAL SERVICES AND THE TOURISM INDUSTRY (PRATO)

As a public administration, Prato is often contacted by commercial subjects (e.g., Google, credit card companies, mobile operators) interested in selling their customers' data to the city. Such an approach is not only expensive for the administration, but it's even questionable from the privacy point of view and sometimes the supplied data is limited to specific context (localisation, payments) and no third parties owning personal data can be included. This introduces limitations in the administration's capability of leveraging citizens' personal data to analyse and improve its service and the same problem can be registered for other parties operating in the city area, like for example cultural institutions (around 20, including theatres and museums).

The DataVaults approach would allow the administration (and other potentially interested parties) to safeguard the users' privacy, while also collecting data from third parties provided by the users themselves, in a safe and privacy-oriented approach, since data will always be in the individual's full control. In addition, the money currently spent to acquire data from operators like Google or similar, could be partially transferred to the citizen data owners through appropriate compensation mechanisms.

3.5.1 Target Audience Reached during Final phase.

3.5.1.1 *Data Owners*

In the final version of the demonstrator, a group of approximately 25 people have been enrolled as data owners, with the aim of testing all the available functions of the DataVaults tools. The group has been formed on the basis of personal contacts managed by the Municipality of Prato and all participants were skilled in ICT.

3.5.1.2 *Data Seekers*

Trials related with the collection of datasets related with questionnaires and routes for both the mobility and culture scenario were carried out in co-operation with the Mobility Office and the Palazzo Pretorio Museum.

In addition, two Data Seekers have been involved for the certificate pilot: a collaborator of the fiscal support centre and representatives of the enrolment office at the site of the University of Florence in Prato.

3.5.2 PRATO Demonstrator Scenarios Evolution

3.5.2.1 *Scenario A - Access to personal data for the analysis of mobility solutions*

3.5.2.1.1 Scenario Description

In this scenario, the Mobility Office acts as a Data Seeker and access the DataVaults platform to look for citizens' personal data (location, itineraries, means of transportation) in order to accomplish different types of activities: a) too plan and verify mobility solutions in the city, also as an integration to the mobility plan released by the administration on a ten-years basis,

and b) to identify adequate samples of citizens for the sending of surveys and questionnaires on traffic and mobility, according to their mobility preference and itineraries, in order to enrich the office knowledge base.

The general approach of the scenario remained still valid, but some adjustments have been introduced, due to the need of synchronising the pilot activities with the technical development of the DataVaults tools, according to the timing of the final available functionalities.

3.5.2.1.2 Objectives of PRATO Scenario A.

Objective	Status	Phase Achieved
To involve a reasonable number of data owners to provide their personal data including mobility behaviours and preferences.	Achieved	Final
To test the platform as a mobility operator (Data Seeker) and check the functionalities of data search and purchase.	Achieved	Final
To evaluate the platform tools for data analytics to improve and integrate the current procedures adopted for mobility planning.	Achieved	Final
To build citizens' samples as a mobility operator, on the basis of specific profiling specifications, and to push them specific questions to enrich the existing knowledge base.	Achieved	Final
To build specific questionnaires and surveys as a mobility operator and send them to the selected citizen samples.	Achieved	Final
To get back sample's answers as a shared dataset according to the data sharing procedures implemented in the platform, including compensation schemas, and analyse them with the platform tools.	Achieved	Final
To provide feedbacks and comments for the technical improvement of the DataVaults app and platform.	Achieved	Beta and Final

Table 23: PRATO Scenario A: objectives

3.5.2.1.3 PRATO Scenario A: Evolution Plan

The scenario objective was that of testing the possibility of the inclusion of citizens' personal data inside the current methodology adopted by the Municipality Mobility Office to design the city mobility plan. In particular, the focus has been posed on the analysis of journey routes recorded and shared by the user group through the DataVaults mobile app and PersonalApp. The analysis of such data has been carried out with the DataVaults SEAS platform.

The scenario deployment has been carried out according to the following steps.

User involvement

A group of approximately 25 users have been involved as data owners, all skilled in ICT: an online training session has been carried out to explain how to use the DataVaults tools, in particular the PersonalApp to collect and share datasets and questionnaire answers and the mobile app to register personal routes. Not all users have been able to use the mobile app, since currently only the version for Android is available.

Sending of the questionnaire

A specific questionnaire of 10 questions has been prepared, on the basis of inputs provided by the Mobility Office as valuable information required in the preparation of the city mobility plan. The questions regarded the preferred mobility means in different life situations, such as work, leisure and shopping, and the problems encountered while using each different transportation means. Some of the questions allowed multiple answers.

The users received the notification from the Municipality of Prato (Data Seeker) asking to answer the questionnaire and provided answers through the PersonalApp interface. They accepted to share the answers and the Data Seeker could then receive them on the DataVaults platform.

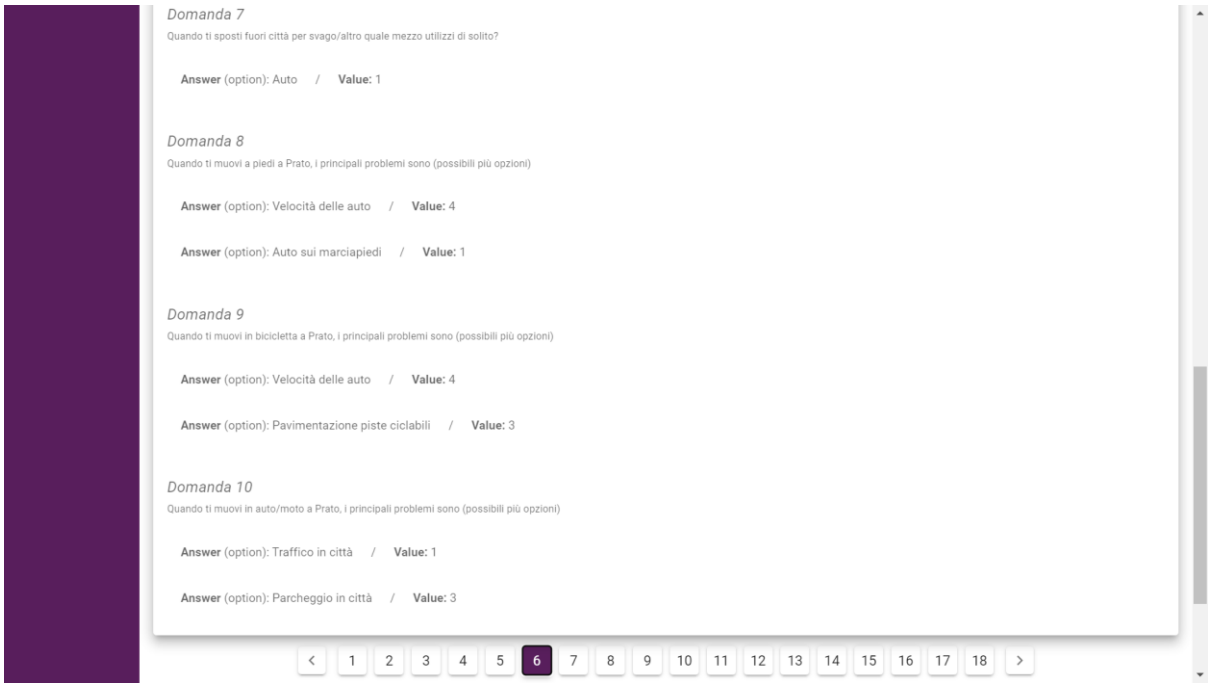

Figure 61: Example of answers provided by the data owners.

Use of the mobile app

Some of the users have been able to record one or more routes by using the DataVaults mobile app on their smartphone. The saved routes have been made available on their PersonalApp in the specific source Route and the users could share them according to their selected sharing policy.

The Municipality of Prato (Data Seeker) was then able to buy such routes through the functionality offered by the DataVaults platform.

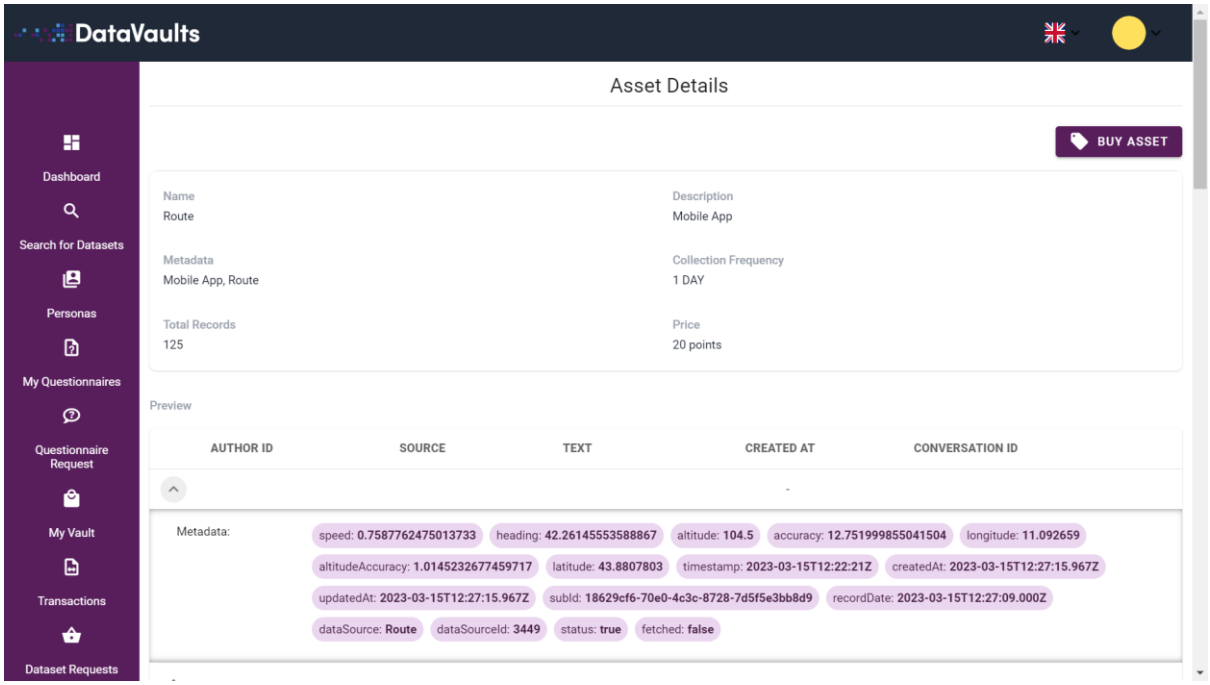

Figure 62: Example of a dataset purchase from the source Route

Analysis of questionnaires and routes

The SEAS tool has been used to make some analysis of both questionnaires and routes collected through the mobile app. In particular, a specific Python code was produced to build bar graphs for the all the answers provided by the data owners to the questionnaire, including the multiple answer option, and show them in the Jupyter notebook.

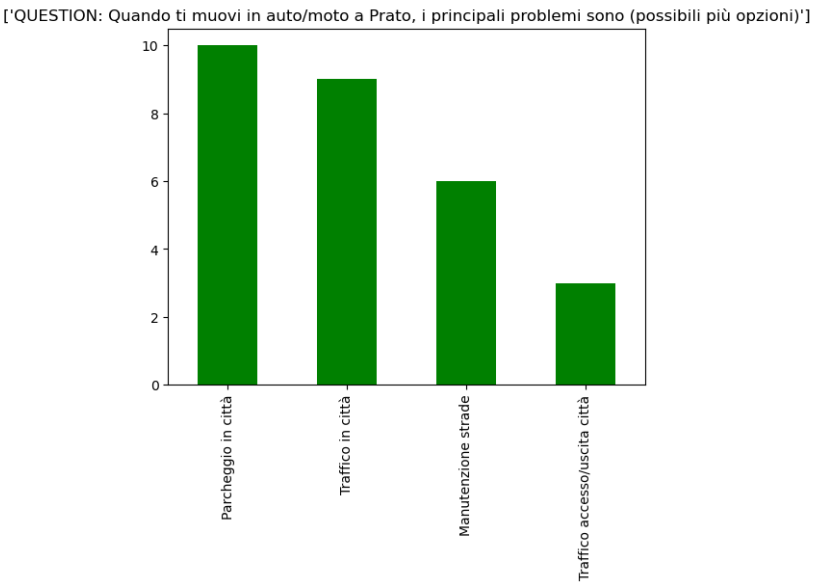

Figure 63: Main problems when moving by car/motorcycle in the city.

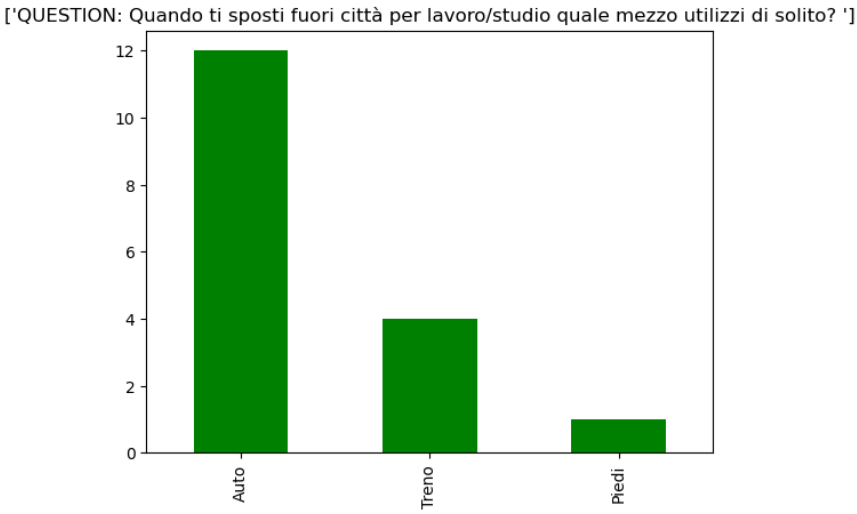

Figure 64: Transportation means, when going out of the city

On the side of the route datasets coming from the mobile app, a specific Python code was produced to map them on a geographic layer and show the aggregation of paths in some specific areas of the city. Datasets were also filtered on different time intervals, to get a picture of user movements in different hours of the day.

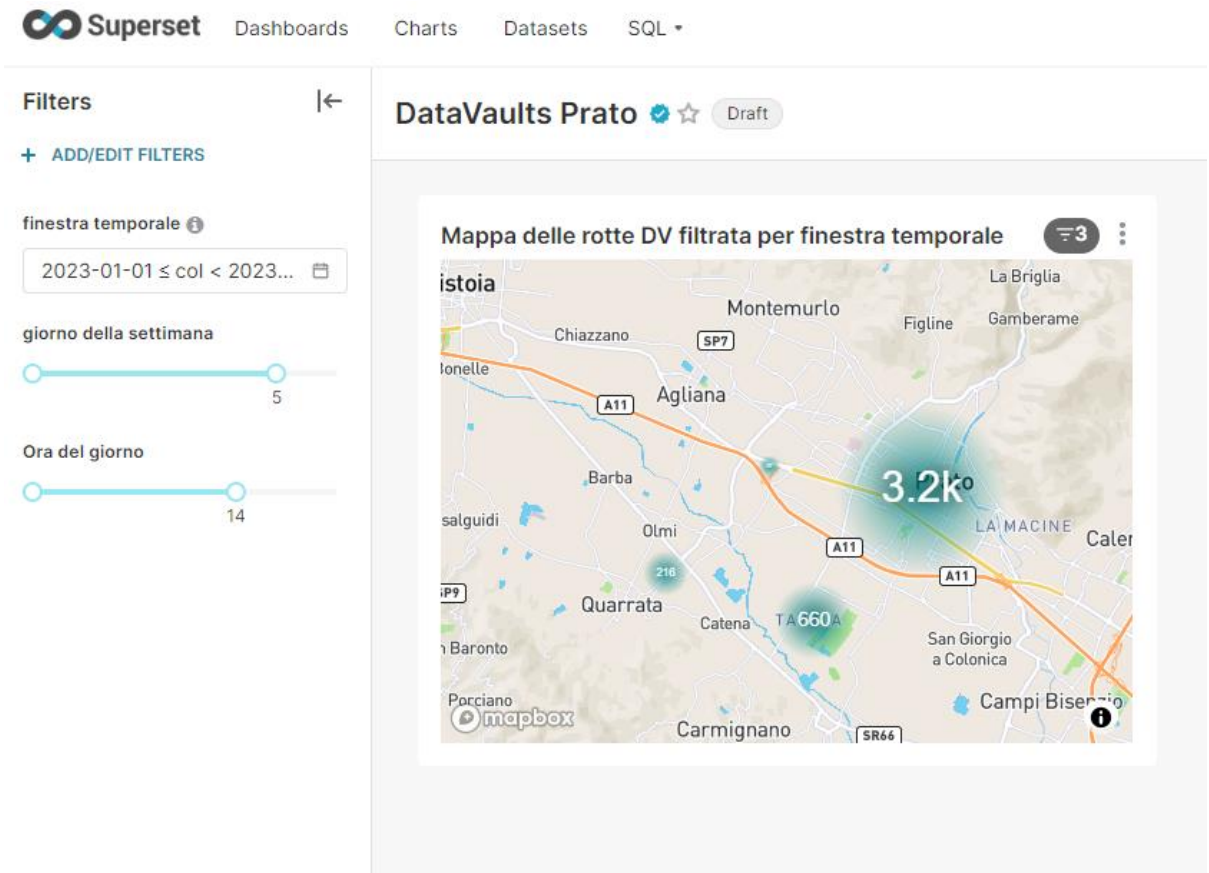

Figure 65: Example of route mapping with filter on day of the week and time

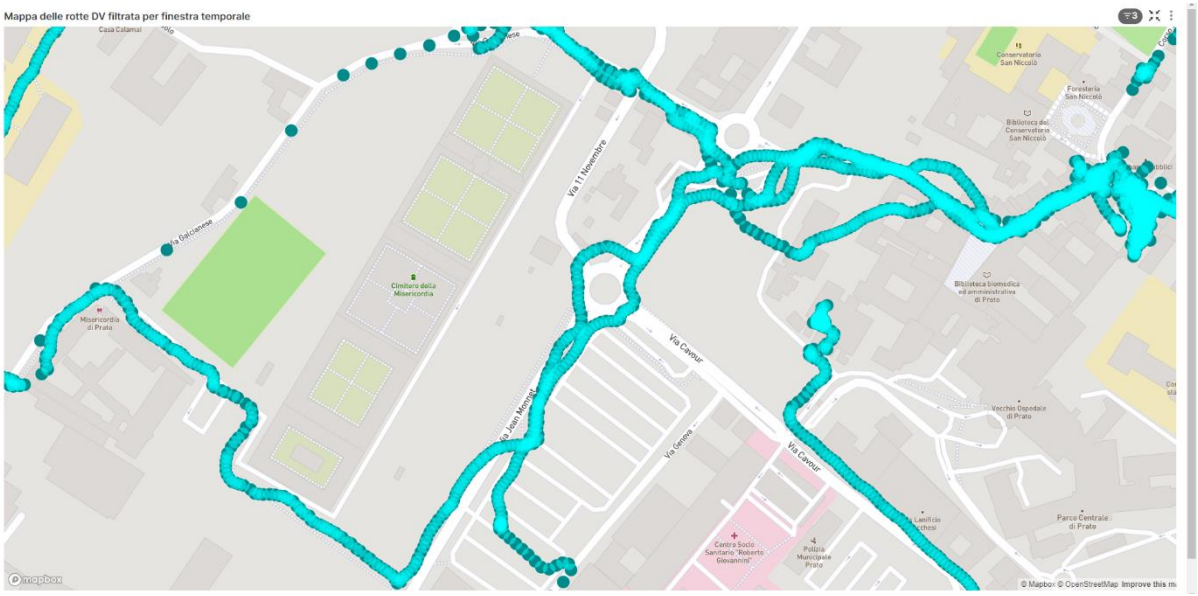

Figure 66: Example of route mapping showing traffic in the city.

The pilot enabled the collection of data from questionnaires that are at the moment collected in more complex and expensive ways. In particular, the prevalent use of the car as a transportation means has been confirmed, together with the problem of parking: the two issues are connected, and it is therefore important to reinforce the work that the administration is already carrying out to improve the use of public transportation and bike lanes in the city.

As far as routes are concerned, it was observed as the main number of passages is concentrated in the city of Prato, which acts as an attraction point for the surrounding communities. In the area where the ICT office is located, the junction near the old hospital is confirmed to be one of the most crowded. In such area, the administration

3.5.2.2 *PRATO Scenario B - Access to personal data for the improvement of cultural offer in the city*

3.5.2.2.1 Scenario Description

In this scenario, a Cultural Institution in the city acts as a Data Seeker and can access the DataVaults platform to look for citizens'/visitors' personal data to accomplish different types of activities, to a) carry out data analysis to improve its cultural offer, and b) define adequate samples of citizens/visitors for the sending of surveys/market campaigns. This scenario needed to be adapted according to the final functionalities provided by the DataVaults tools. In particular, it was decided to focus the scenario objectives on the questionnaire delivery and management to facilitate the collection of information useful for the Palazzo Pretorio Museum to better evaluate citizens' cultural interests.

3.5.2.2.2 Objectives of PRATO Scenario B

Objective	Status	Phase Achieve
To involve a reasonable number of data owners to provide their personal data including cultural preferences, attendance and liking of cultural events.	Achieved	Final

To evaluate the platform tools for data analytics to improve and integrate the current procedures for the planning of cultural offer.	Completed	Final
To build specific questionnaires and send them to the selected citizens' samples.	Achieved	Final
To get back sample's answers and analyse them with the platform tools.	Achieved	Final
To provide feedbacks and comments for the technical improvement of the DataVaults app and platform.	Achieved	Beta and Final

Table 24: PRATO Scenario B objectives

3.5.2.2.3 PRATO Scenario B: Evolution Plan

The scenario objective was focussed on the collection and analysis of information useful to cultural institutions to better know their visitors' interests and improve their marketing activities, through the sending of questionnaires managed with the DataVaults tools.

User involvement

The user group was the same as scenario A: approximately 25 users, all skilled in ICT. They all attended the online training session organised to explain how to use the DataVaults tools.

Sending of the questionnaire

A specific questionnaire of 6 questions was prepared, based on inputs provided by the Museum of Palazzo Pretorio: the questions regarded the modalities of the visit to the museum, the evaluation of the quality of such experience and possible visits to other museums of the city.

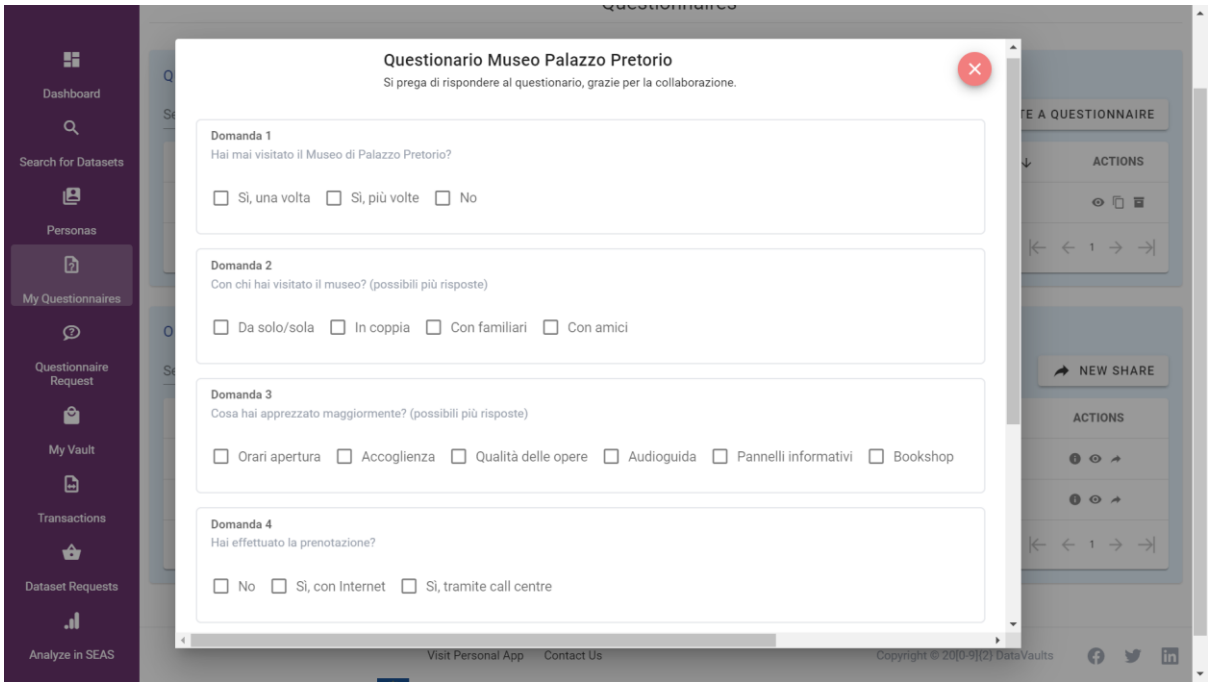

Figure 67: Questionnaire delivered to the data owner from the museum Data Seeker

The users received the notification from the Municipality of Prato (Data Seeker), asking to answer the questionnaire, and provided their answers through the PersonalApp interface.

They accepted to share the answers and the Data Seeker could then receive them on the DataVaults platform.

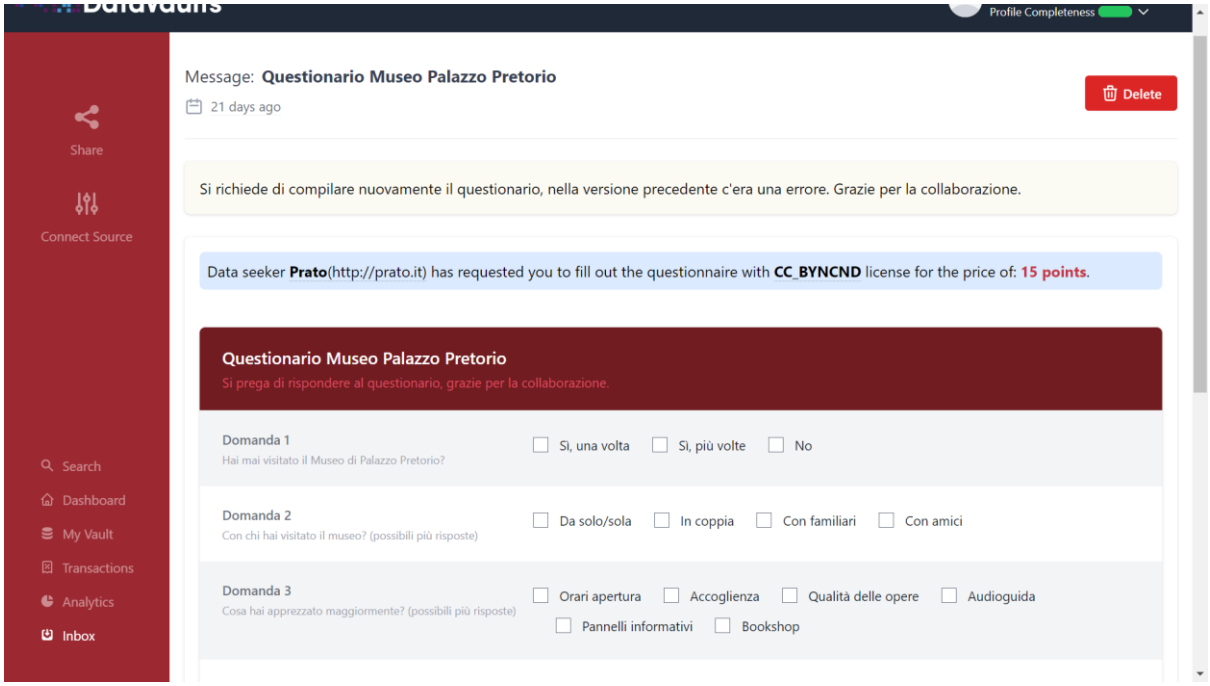

Figure 68: Data owner receiving the questionnaire from the museum Data Seeker

Analysis of questionnaires

The SEAS tool has been used to make some analysis of the returned questionnaires in terms of bar graphs reporting the summary for the all the answers provided by the data owners, including the multiple answer option for the graphs were shown in the Jupyter notebook.

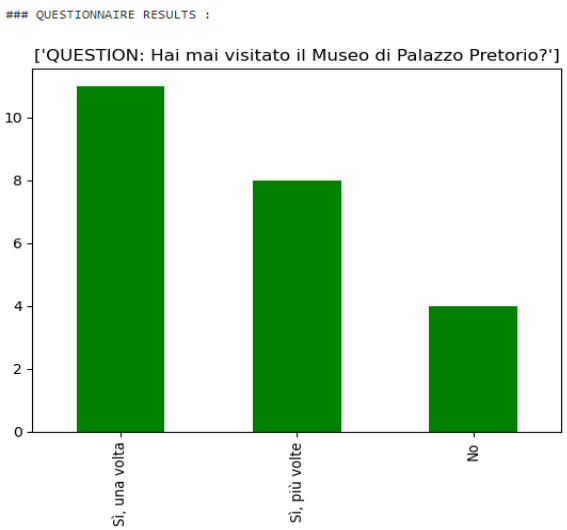

Figure 69: Number of visits to the museum

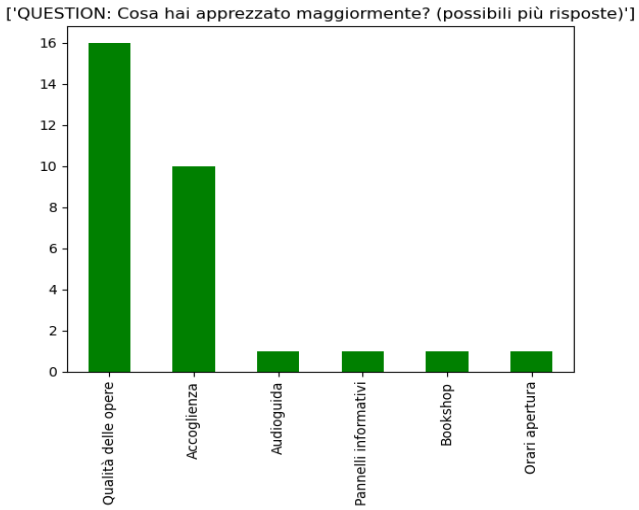

Figure 70: Most appreciated aspects during the visit to the museum

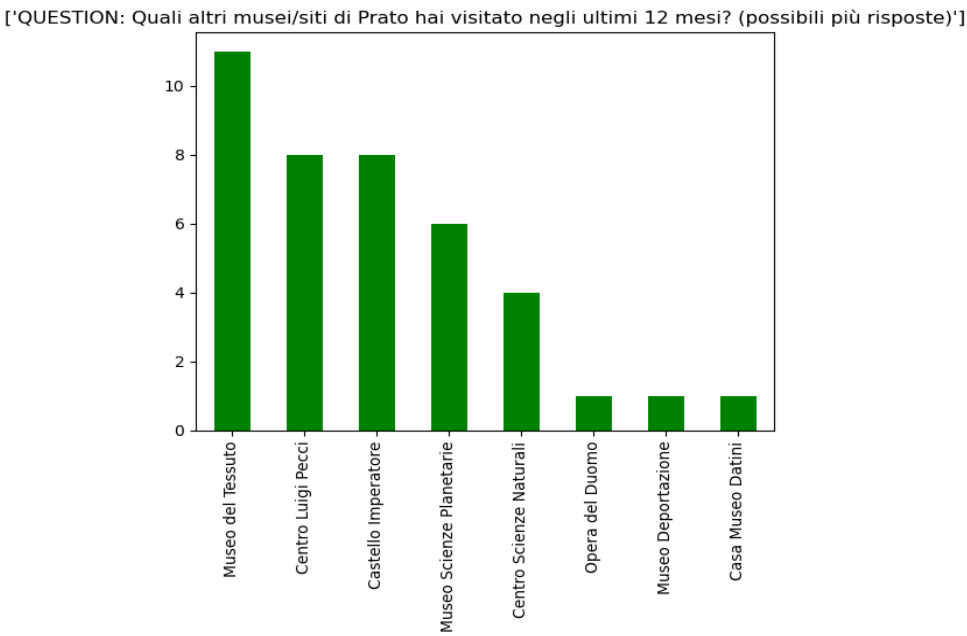

Figure 71: Visits to the other museums in the city

It was interesting to verify that several users have visited the Palazzo Pretorio Museum more than once and that, beside the quality of the exhibition, also the quality of the welcome has been appreciated by the users. There is also much interest in other museums of the city and, maybe a little surprisingly, the Planetarium looks one of the most visited.

3.5.2.3 *PRATO Scenario C - Access to personal data for the delivery of personal certificates*

3.5.2.3.1 Scenario Description

In this scenario, the CGIL-CAAF fiscal support centre accesses the DataVaults platform as a data seeker, with the objective of acquiring personal civil certificates from a user requiring some fiscal services. The main objective was to test a new procedure for the exchange of personal certificates between a citizen and a third party requiring such certificates to provide

specific services, in order to ensure the citizen's full privacy in such exchange and to make it more in line with the EIDAS digital wallet approach, now fostered by the European Commission.

3.5.2.3.2 Objectives of PRATO Scenario C

Objective	Status	Phase Achieved
To verify the software connection (API) between the city population registry and the DataVaults App and implement the inclusion of the civil certificates inside the project data model.	Achieved	Beta
To test the proposed innovative document exchange between citizens and the CGIL – CAAF fiscal support centre by involving a small group of users (2-3 citizens).	Achieved	Final
To evaluate the whole procedure to check strong and weak points and suggest possible technical improvements, from both data owner's and data seeker's point of view.	Achieved	Final

Table 25: Prato Scenario C: objectives

3.5.2.3.3 PRATO Scenario C: Evolution Plan

The scenario aimed at testing the DataVaults approach in the management of personal certificates, in order to facilitate the collection of a citizen's own documents in a safe place (DataVaults PersonalApp) and allow the citizen to share them autonomously with third parties requiring such certificates to deliver specific services to them.

Data Owner involvement

The same user group as in Scenario A and B was trained also on how to use the DataVaults PersonalApp to collect and share their personal certificates from the population registry of the Municipality of Prato by means of the DataVaults tools.

Data Seeker involvement

Two Data Seekers have been involved to test the whole flow for recovering user's personal certificates: a collaborator of the fiscal support centre and the enrolment office at the site in Prato of the University of Florence.

Sharing and download of personal certificates.

The Data Owners entered the DataVaults PersonalApp and connected to the Prato source to collect their own certificates through a specific API developed as a functionality of the DataVaults platform: the API retrieves the Data Owner's fiscal code and his/her personal certificates from the population registry of the Municipality of Prato.

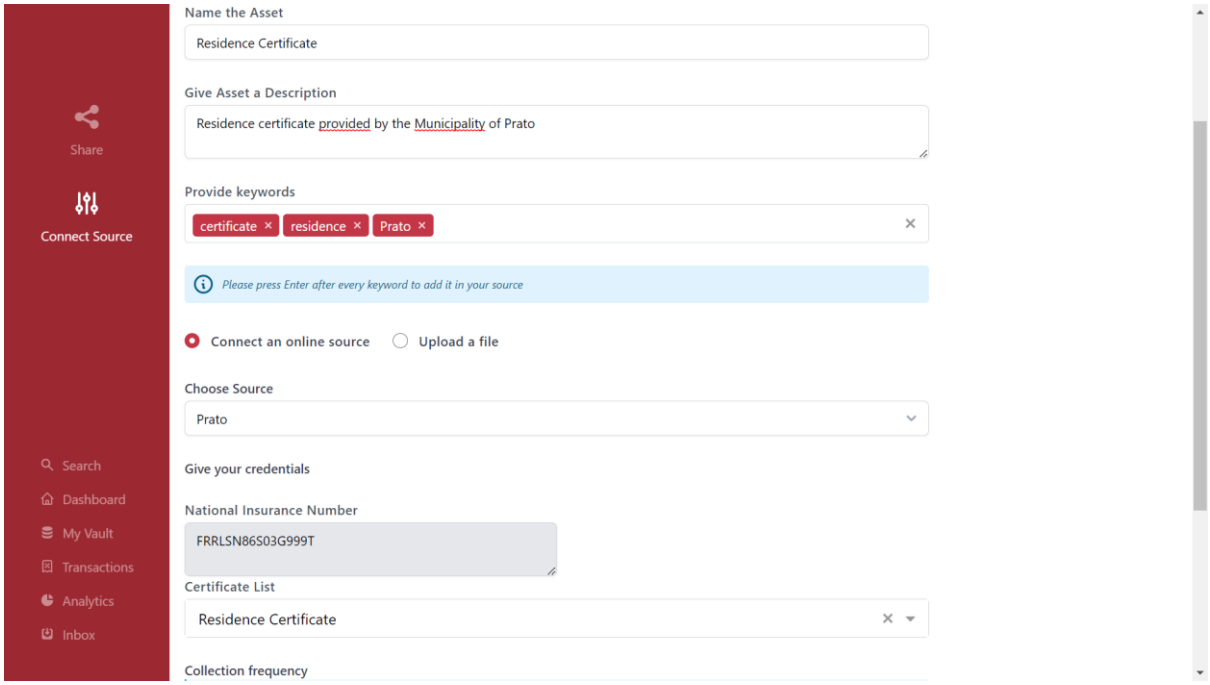

Figure 72: Data Owner collecting own certificate from the Prato population registry.

The Data Seeker then entered the DataVaults platform to search for personal certificates by the user's fiscal code, which is the usual search key element in their procedures. He/she completed the whole procedure and was able to download the requested certificate.

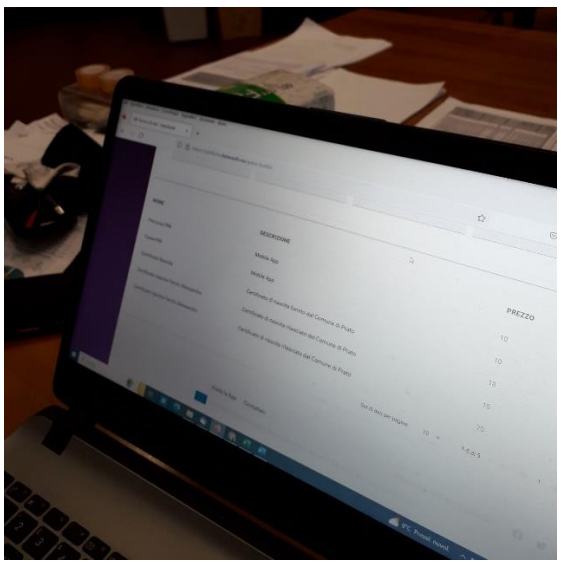

Figure 73 :Search of certificates on the DataVaults platform

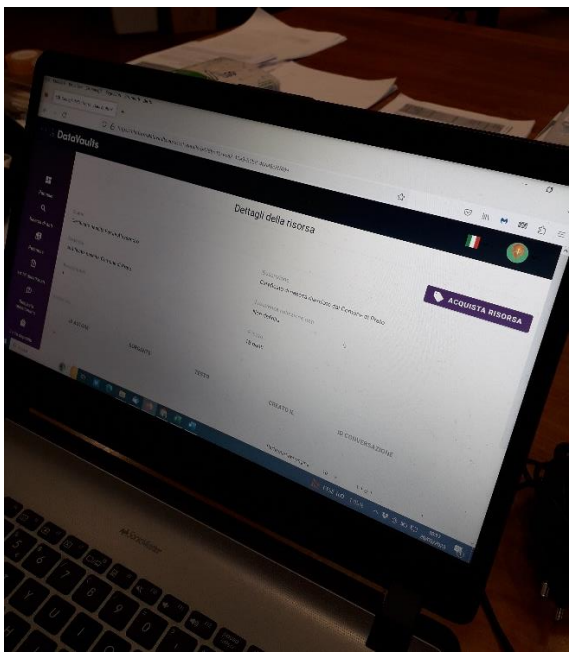

Figure 74: Data Seeker collecting Data Owner's certificate.

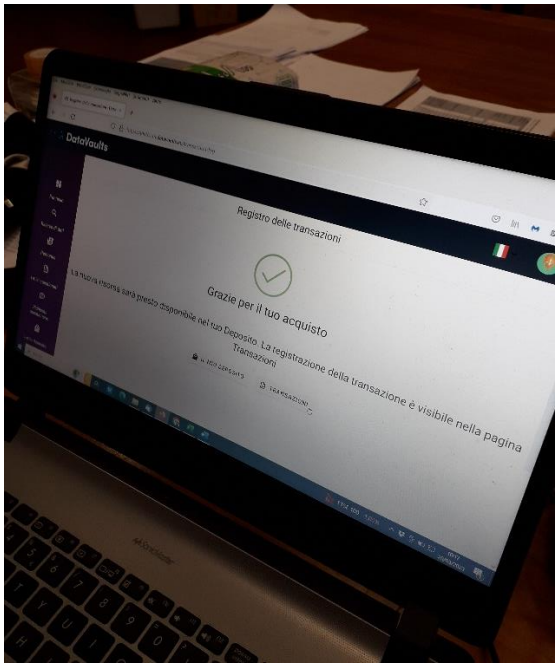

Figure 75: DataVaults platform confirming certificate transaction.

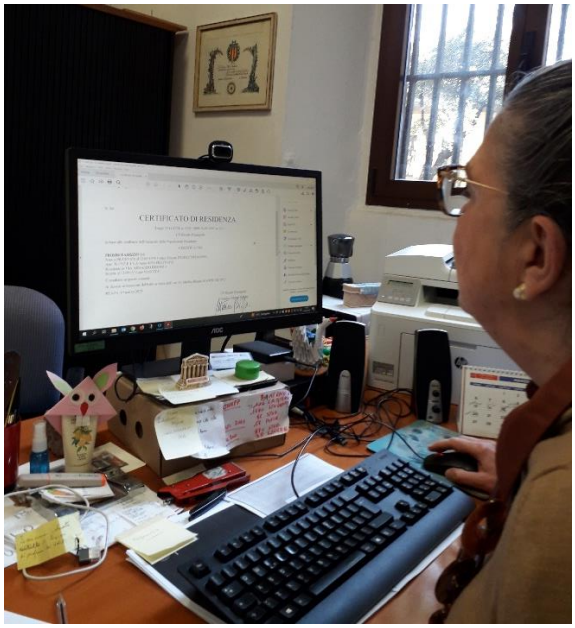

Figure 76: Data Seeker downloading the certificate from the DataVaults platform

The experience in the pilot was satisfactory and the scenario confirmed the possibility to adopt the DataVaults platform also as a valuable tool for the exchange of personal documents between Data Owners and third parties, in full respect of the user's privacy.

3.5.3 High Level Impact Recorded in the Demonstrator

3.5.3.1.1 Benefits for the Data Seekers recorded from the Scenario.

As far as scenario A on mobility is concerned, the DataVaults tools can allow to obtain, in a quick and relatively simple way, the map of city areas where traffic is more concentrated. Currently, the administration gets such information by means of specific and expensive sensors, often put in place for a limited period of time, while the DataVaults approach might allow to get data in a more continuous and rather simple way, even by profiling users and reducing costs for the administration in planning new solutions for the city mobility.

As far as scenario B is concerned, the modality of sending questionnaires tested with DataVaults allows to identify specific targets, without the need for specific registration on the museum website. Currently, this activity is carried out by external companies that are usually rather expensive and this is the reason why such investigations are made not very often, while the DataVaults approach can be very useful for the setup of more continuous surveys and monitoring of results.

On the whole, we can say that, with respect to what was reached in the pilots, the effectiveness of the approach proposed by DataVaults would be increased as far as a relevant number of data source is available, so to enrich the information framework and the opportunities for both data owners and data seekers.

As far as scenario C is concerned, the DataVaults approach might provide a very interesting opportunities for data seekers requiring personal documents, without the need of the data owner's physical presence and even without the need to alert the data owner of such request, since the platform guarantees the full respect of privacy and permission requirements. To this extent, DataVaults looks also in line with the EIDAS digital wallet that is under development.

3.5.3.1.2 Benefits for the Individuals recorded from the Scenario.

Individuals as data owners can provide their personal data to the administration and to third parties in an easy and protected way, in full respect of their privacy, and this can also represent a positive incentive for citizens to cooperate with the administration in the design and management of urban policies, particularly from the smart city point of view.

From the operational point of view, it must be added that, despite the relevant effort of the technical project team to develop easy user interfaces, data owners had some difficulties in managing the system functionalities and in understanding the different aspects connected with the privacy issues. In fact, the current attitude on users' side is rather to ignore this topic and just use online services offered by the big players (Facebook, Twitter, Amazon, etc.), without realising the implications on their privacy.

3.5.3.1.3 Benefits for the Organisation recorded from the Scenario.

From the public administration's perspective, there is a clear advantage in the possibility of accessing citizens' personal data without being obliged to buy such data from big players, as it's now the case, with huge costs and no guarantee on privacy respect.

In addition, the opportunity to offer DataVaults as a service platform to the citizens, may allow the public administration to contribute in increasing public awareness on the privacy issues: by establishing a contract between PA and citizens, city governments could play a key role in addressing power unbalances of the current data landscape, acting as trusted data intermediaries and enabling the use of citizens' personal data for the public interest.

A final consideration can be made on compensation: although the DataVaults approach recognises the possibility of rewarding citizens for their sharing of personal data, during the pilot it was acknowledged that this aspect is very difficult to manage for public administrations, due to regulatory constraints that prevent the recognition of monetary and non-monetary benefits to citizens. This aspect should then be managed through a separate channel, like for example an online shop offering rewards managed by private business.

3.5.3.1.4 Demonstrator KPIs

In this section, we present the KPIs relevant to the demonstrator as identified in deliverables D6.1 and D6.2, as well as other KPIs recorded during the demonstration phase.

It is noted, that not all these KPIs could be measured during the demonstrator phase of the project, as some of them concern post-project targets, as part of the exploitation activities of the demonstrator. Those are marked as such "Long-Term Impact" KPIs and are discussed in deliverable D7.4

Objective	Metrics: Description of indicators towards assessing progress	Measures of change: Success criteria	Data Collection Methods and sources	Status at M40	Comments
Scenario 1 - Mobility					
To increase the number of data owners involved	Currently zero data-owners involved	To increase to 100.	Number of registrations on the app/platform Check of registration file	30	Reshaping of pilot actions in accordance with technical development
To increase the number of available data sources	Currently zero	5	To be extracted from app functionalities. Check of App APIs	2	Due to technical revision some sources were not included
To increase the number of shared datasets	Currently zero	200	To be extracted from the platform data base	70	Reshaping of pilot actions in accordance with technical development
To increase the number of activated smart contract	Currently zero	100	To be extracted from the platform distributed ledger	70	Due to technical revision some sources were not included
To increase the number of data analysis procedures	Currently zero	3	To be extracted from the data seeker workflow. Production of data analysis report.	2	
To increase the number of questionnaires/surveys	Currently zero	2	To be extracted from the data seeker's workflow on the platform- Data seeker's documents.	1	
Improvement in the planning capabilities as perceived by the Office	Survey (Likert scale 1-5)	N/A	Interview	4	
Savings in the installation of traffic sensors and data acquisition procedures	Specific budget from the Mobility Office's records	10.000 euro	-50% Survey results	The savings could be up to 100% in case a consistent number of users is involved	

Scenario 2 - Culture					
To increase the number of data owners involved	Currently zero	50	Number of registrations on the app/platform. Check of registration file.	30	Reshaping of pilot actions in accordance with technical development
To increase the number of available data source	Currently zero	5	To be extracted from app functionalities. Check of App APIs	1	Due to technical revision some sources were not included
To increase the number of shared datasets	Currently zero	100	To be extracted from the platform data base	30	Due to technical revision some sources were not included
To increase the number of activated smart contract	Currently zero	50	To be extracted from the platform distributed ledger	0	only questionnaires used
To increase the number of data analysis procedures	Currently zero	3	To be extracted from the data seeker workflow. Production of data analysis report	1	
To increase the number of questionnaires/surveys	Currently zero	2	To be extracted from the data seeker's workflow on the platform. Data seeker's documents.	1	
Improvement in the planning capabilities of the cultural institutions	Survey (Likert scale 1-5)	N/A	Interview	4	
Savings in data acquisition and analysis procedures	Specific budget figure	5.000 euro	-50%	No specific reduction can currently be established as the DataVaults tools work in addition to other tools	
Scenario 3 - Certificate					
To increase the number of involved data owners	Currently zero	10	Number of registrations on the app/platform.	30	

			Check of registration file		
To increase the number of shared datasets	Currently zero	10	To be extracted from the platform data base. Check of the data owners' activity	30	
To increase the level of Prato administration's savings of resources in terms of costs and personnel	Cost of one certificate release operation at the counter [15 min personnel cost + paper cost] Current cost €4.10	~ €0	n. data owner x current cost at the counter	~ €0	
Scenario Wide KPIs					
To increase the level of Data owners' satisfaction in using the DataVaults tools	N/A	4.5	Survey results (Likert scale 1-5)	3.1	
To increase the level of Data seekers' satisfaction in using the DataVaults tools	N/A	4.5	Survey results (Likert scale 1-5)	4.1	

Table 26: Demonstrator #5 KPIs

3.5.4 PRATO Demonstrator's Activities Timeline

The following table presents the main activities performed within the demonstrator.

Demonstrator5 / PRATO	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	M37	M38	M39	M40
Scenario A - Access to personal data for the analysis of mobility solutions																						
Alpha Phase																						
Check of tool development																						
Set up of a small testing group																						
Test activity on the platform basic functionalities																						
Collection of feedbacks																						
Expanding the pilot																						
Set up of a larger user group																						
Test of advanced platform functionalities																						
Planning and testing of questionnaire preparation and sending																						
Planning and testing of data analysis procedures																						
Evaluation of results																						
Consolidating the pilot																						
Extension of the user group																						
Data collection, questionnaire and data analytics in a real context																						
Data analysis for mobility planning and evaluation of results																						
Scenario B - Access to personal data for the improvement of cultural offer in the city																						
Alpha Phase																						
Check of tool development																						
Set up of a small testing group																						
Test activity on the platform basic functionalities																						
Collection of feedbacks																						
Expanding the pilot																						
Continuous set up of a larger user group																						
Test of advanced platform functionalities																						
Planning and testing of questionnaire preparation and sending																						
Planning and testing of data analysis procedures																						
Evaluation of results																						
Consolidating the pilot																						
Continuous extension of the user group																						
Data collection, questionnaire, and data analytics in a real context																						
Data analysis for the planning of cultural events and evaluation of results																						
Scenario C - Access to personal data for the delivery of personal certificates																						
Testing the procedure																						
Release of the API with the population registry																						
Test of the certificate flow procedure																						
Collection of feedbacks																						
Consolidating the procedure																						
Set up of a user group																						
Verification of the certificate sharing procedure in a real context																						
Result evaluation																						

Table 27: Execution Timeline for Demonstrator #5 – PRATO

4 COMMUNICATION/INTERACTION WITH STAKEHOLDERS

4.1 STAKEHOLDER ROLES

When the evaluation framework was set out in Chapter 6 of D6.1, the starting point in it was the identification of all the stakeholders of significance to DataVaults at the outset.

And within WP5, Task 5.1-DataVaults Platform Requirement and User Stories Elicitation is devoted to systematically aggregate and analyse the user requirements of all stakeholders involved in the DataVaults value chain, to conclude on the requirements that should be met by the DataVaults platform.

Similarly, the exploitation and dissemination WPs contributed to the analysis as to who the DataVaults stakeholders were.

D6.4 presented the evaluation of the state of readiness at M24, which was concluded to be satisfactory and having met all the expectations set, including how we had interacted with and responded to all the stakeholders which had been identified in D6.1 within the Evaluation Framework. All the positive activities reported in D6.3 and D6.4 have since been continued and consolidated.

Positive responses can be made in respect of the questions which we first identified in relation to dealing with the DataVaults stakeholders.

- We identified the stakeholder's roles in evaluation planning, implementation, interpretation of results and the decision-making around any next steps.
- We kept reviewing the list of identified stakeholders to ensure all appropriate stakeholders were included.
- We created a plan for continuing stakeholder involvement and a communication strategy.
- Relevant areas were identified for stakeholder input.
- Stakeholders have been brought together as needed.
- Key stakeholders have been targeted for regular participation.
- We have involved stakeholders in the evaluation process and the online engagement tracker was constantly revised by all partners.

Also, the carrying out of the activities set out in the Stakeholder Engagement Plan within D8.1 has provided a solid mechanism for further interactions to occur and assist in the shaping of the project as it progresses.

4.2 THEORY OF CHANGE: STAKEHOLDER PARTICIPATION AND COOPERATION

The use of the Theory of Change at the evaluation stage can verify whether the project had included a satisfactory approach for sharing information and encouraging cooperation with partners, national/local project stakeholders and other EU projects and programmes with the demonstrations completed and the project in its final phases. In using the Theory of Change

at the evaluation stage, stakeholder analysis assisted in the identification of the key stakeholders along with their respective roles, capabilities, and motivations in each step of the causal pathways from activities to achievement of outputs, direct outcomes and intermediate states towards impact.

Will we have engaged with the eco systems we are aligned with and mirroring the current strategic activity from the Commission, for example? Chapter 6 reports in detail on the success achieved in these aspects and is further evidenced in D7.4.

During the main evaluation process the initial Theory of Change at the design stage was discussed with the key actors and stakeholders involved in the implementation of the DataVaults project. Revisions and updates to the Theory of Change were made to reflect any changes in the project's intended results or intervention logic and to take into account any changes in external context of the intervention that may influence the causal pathways and the changing needs and priorities of stakeholders.

A clear example of this adaption to the needs of identified stakeholders can be seen in the recognition that the majority of the DataVaults partners had referred to their business interest in exploring smart cities as an avenue for exploitation and this became a focal point.

Similarly, the project responded regularly to the changes in emphasis within EC policy and strategy. For example, the Smart Cities MarketPlace initiative led by DataVaults has subsequently heeded requests to re-focus on feeding into the evolving Smart and Sustainable Cities DataSpace, following on from advice to collaborate with Livingin.eu and work alongside Open and Agile Smart Cities in fostering adoption of the Minimum Interoperability Mechanisms.

4.3 STAKEHOLDERS AND DATAVAULTS

Work in WP7 and in D8.1 Dissemination, Communication and Stakeholder Engagement Plan set out a methodology which DataVaults would embrace to engage with stakeholders, which would be shared with the exploitation managers in WP7, having their own relationships with stakeholders. Specifically, Task 8.4-Stakeholders' Outreach, Liaisons and Engagement Activities established that in addition to the standard dissemination and communication activities, the DataVaults consortium would additionally be engaged in dedicated stakeholders clustering and in decision makers engagement and awareness raising.

These activities targeted mainly the decision-making communities, namely the communities that hold more potential in commercially exploiting the results and applying them in daily practice and in regional or even national extent. Capitalising on already formulated ecosystems, there were pro-active measures taken to enhance them with additional stakeholder groups, such as policy formulators, decisions makers, policy makers, organisations and initiatives focused on entrepreneurship, business support services, etc.

It was natural during the project for partners also to engage with the stakeholders which were of most relevance to them, especially for the five demonstration partners. Each had well-defined interactions with their own stakeholders. The DataVaults individual partner's

exploitation plans and potential scenarios based upon them set the lead in identifying the stakeholders which are most likely to be of most value.

Additionally, Task 1.3 stated that "The aim will be to *indicate all stakeholders' points of view*, containing both functional and non-functional requirements, and to generate a high-level description of the expected behaviour of all sub-systems that are going to be specified and developed." Scenarios were generated with a view to returning to them through the project. What was born in mind, whilst creating these high-level scenarios, was the logic behind them, which is that they were expected to contribute to the overall success of the project.

The Table in Appendix 2 reflects success in delivering the requirements which were captured from stakeholders within that process.

4.4 MAIN STAKEHOLDERS ENGAGED WITH

Engagement with stakeholders during the demonstration process consisted of dealing with a wide range of stakeholders grouped into 16 categories in D6.1. The main groups of stakeholders we needed to engage with included the following:

4.4.1 Primary Personal Data Providers (Individuals)

This included all the individuals which were generating and collecting their personal data from various services, devices, and applications. It was these data which was considered "personal" and constituted the core data of that was of interest to the DataVaults project. As well as the core function of supplying personal data, they act within the demonstrators in a variety of roles, as citizens, supporters and customers:

- Travelling
- Being tourists
- Attending sports and leisure events,
- Being athletes
- Establishing businesses
- Consuming energy
- Interacting with health services

Each demonstrator has reported on the numbers and types of citizens they have engaged with and their plans for such engagement in the future.

4.4.2 Economic Operators.

These were data seekers (also titled as 1st-tier economic operators), that looked for enjoying business intelligence based on Primary Personal Data. In this tier, data seekers (organisations of any type) can work on the data of the first tier (primary data) and combine them with other types of data they have in order to create new datasets or relevant derivatives (insights, reports, etc.). D7.4 lays out the scenarios envisaged for taking this aspect forward at scale, with individual demonstration sites having started the process. Other stakeholders will have the opportunity to provide services based on data re-used and re-sold. D7.4 illustrates progress in moving forward on this front.

4.4.3 The European Commission

A primary stakeholder is the European Commission. The project has legal requirements regarding the contract with the Commission, which needed to be fulfilled in the high-level scenarios and which will be covered in Chapter 6.

4.4.4 Project affiliations

The project list was expanded substantially as the project progressed, and whilst the WP8 Dissemination Tracker updates the list of projects, it does it from a dissemination perspective. However, the key focus for the deep engagement with projects has been to enhance the likelihood of exploitation for DataVaults and the close engagement with 16 projects in the production of the DataVaults-led book yielded great potential for further collaboration, but also countless lessons learned of real value to future deployments of DataVaults. Indeed, interactions with such stakeholders as led to further exploitation opportunities arising.

4.5 ROLES AND REQUIREMENTS OF STAKEHOLDERS

Understanding the roles and requirements of the different categories of stakeholders made it easier to attempt to evaluate whether we were successful in reaching them during the demonstrations.

Whilst we cannot claim to provide solutions answering all the requirements referred to below during a single project, it is worthwhile illuminating them as indicators of the direction we need to travel to sustain and exploit DataVaults.

Stakeholder group	**Their requirements**	**Progress towards achieving them**
Individuals	Improved control and awareness of how their data are shared and managed.	Fulfilled
	Remuneration based on the data produced and shared.	Fulfilled
	Better service provision through use of their data.	Fulfilled
Data Industries	Easier and seamless access to personal data, using secure and privacy aware guarantees.	Fulfilled
	Significantly increased opportunities related to integrated data and data integration services' provision.	Fulfilled
	More evidence-based analytics to support their strategic and operational decisions	Fulfilled
	Innovative and more effective products and services.	Fulfilled
	Significantly reduced time to market for new products and services.	Conditions fulfilled to do this
Entrepreneurs	Significantly increased business opportunities related to innovative services and apps.	Fulfilled
	Easier and seamless access to constantly growing volumes of cross-sectorial multilingual big data.	Fulfilled

	New business opportunities related to the building on top of existing solutions	Fulfilled
Data Scientists	Improved and fast access to personal data allowing them to focus on experiments development rather than investing in data management and collection issues that need to be tackled due to regulation.	Fulfilled
	Secure environment for experimentation with sensitive personal data structures.	Fulfilled
Demonstrators	Improvements to the services they currently offer	Fulfilled
	New methods of improving these services	Fulfilled
	Closer relationships with their citizens/customers as stakeholders.	Fulfilled
Policy Makers	Faster and more effective decision-making procedures based on personal data	Fulfilled
	Reference implementation to based future legislation and regulations for personal data	Fulfilled
	Contributions to European, national and local policy	Fulfilled
	Contributions to the standardisation processes	Fulfilled
Society at large	Advancing research and applying innovative technologies that take the best of breed personal data management	Fulfilled

Table 28: Roles and requirements of different stakeholder groups.

4.6 OVERLAP WITH WP8 ACTIVITIES

WP6 activities proceed in tandem with WP8 by coordinating the communication and interaction with the various types of stakeholders. Notably, the dissemination team extends its activities much further than knowledge diffusion and reaches out to key stakeholders from various sources (companies, affiliated projects, clusters and initiatives, associations, other key collaborators etc.) that have the expertise to provide valuable advice and assist in the overall evaluation process through their experience and work in related fields.

The KPIs reported in WP8 deliverables are also interconnected with liaisons, synergies, face-to-face feedback and the two-way interaction from stakeholders, whose importance we have shown in the Evaluation Framework.

The implementation of the dissemination and communication strategic plan included specific activities dedicated to synergies and liaisons that engages stakeholders in the evaluation process and these are reported in D8.4 as well as where they contribute to the exploitation strategy, and these are reported in D7.4. Input was strengthened from the full range of DataVaults stakeholders, through a close collaboration between WP6 and WP8, with WP7 being the clear beneficiary in developing its exploitation plans.

In practical terms, the contributions included the following activities:

1) We have expanded our network reach to more than 300 organisations, companies, external partners of affiliated projects and other entities from the real market, using as basis the pre-existing liaisons, vendor and supplier, client, and collaborator

networks, that the consortium partners had before entering the consortium. We expanded by not only presenting the project and its novel aspects but also had the opportunity to engage through oral discussions about the benefits that DataVaults can present with its innovative approach, as well as receive feedback and ideas for optimal use in the future and potential exploitation paths, after the project's contractual end.

2) 41 distinct EU-Funded Research Projects have been reached and been affiliated under the scheme "sister-projects" which can also be found in the relevant section of our project website. Beyond promotional activities exchange and social media links, we had fruitful discussions for mutual collaboration. These extended to the point of considering whether DataVaults could see the partners of these projects as potential stakeholders, and what benefit could be derived through joint future use at exploitation stages. Their evaluative and constructive comments were noted and taken into account.
3) In the same scope, we managed to engage and interact with 10+ industrial communities, post 12 times in the relevant EC websites (such as CORDIS) with dedicated blogposts and achieved 50+ digital promo actions such as magazines (digital and printed) as well as more than 100+ similar smaller posts in various media.
4) Numerous 3rd party stakeholders (more than 80+) have been reached with fruitful discussions through our 3 main "key liaisons" which have helped throughout the whole duration of DataVaults. Namely: namely Major Cities of Europe, BDVA and OASC (Open and Agile Smart Cities) under the umbrella of EC Smartcities Marketplace.
5) The latter has been the major supporting organisation that has led to our book publication, which was co-ordinated scientifically by WP6 and ASSENTIAN, funded by WP8 UNISYSTEMS, and populated with content by all consortium partners. Within the book all demonstrators contributed with chapters, and the book has enjoyed great success in the recent Barcelona event and elsewhere, drawing attention to many external stakeholders.
6) Towards systematic tracking of stakeholder engagement, the online dissemination & engagement tracker was constantly revised by all partners. Of course, we also took into account the survey run by DataVaults partner ETA and the stakeholder feedback received from 400+ participants.
7) With the assistance of WP8 communication experts and under the guidance of WP6 experts with business driven and evaluation driven scope, we have completed for all Five (5) demonstrators the following:
    a. 5 promotional videos also posted on our DataVaults YouTube Channel
    b. Approximately 10 demonstrator events of stakeholder engagement (which is on average 2 per demonstrator, but every demonstrator had at least one) in which the general public was invited for presentation and quick training on how DataVaults operates and tackles critical issues. These were co-ordinated and run by the demonstrators themselves, under the supervision of WP8 and WP6, and recording was made in the dissemination tracker of all these activities.
    c. In most of these events, an explanatory training-webinar with presentation to stakeholders of the relevant community of each demonstrator, was included.

A brief session of questions-clarifications based on participant enquiries, followed.

d. The demonstrator events were supported by the promotional pack that WP8 prepared, namely, banners, roll-ups, 3 brochure versions, posters, graphics, and other dissemination material available at the project's digital repository. For speed and ease of re-usability, demonstrators printed directly, distributed, and used this material for their stakeholder engagement events, depending on their needs for each event.

8) The above-mentioned demonstrator events (10 totally, instead of 2 initially planned in the Grant agreement), together with the Expo participations in digital booths such as within BDVA events, were implemented to replace the initially planned hackathon that the technical co-ordinators of the project decided to cancel in the last trimester and replace it with all these demonstrator events that were not initially foreseen in the DoA.

Thus, the carrying out of the activities set out in the Evaluation Plan were complemented by those coming out of Stakeholder Engagement Plan within D8.1 which provided a solid mechanism for further interactions to occur and assist in the shaping of the project as it progressed.

5 TECHNICAL ASPECTS – TECHNOLOGY ACCEPTANCE BY THE DEMONSTRATORS

Section 5 is focusing on the technology acceptance of the platform as perceived by the demonstrators, while the technical evaluation of the platform has been performed under Wp5. To be able to measure the quality of the system from a user's perspective (e.g., non-software performance level evaluation) DataVaults makes use of the ISO 25010 Quality in Use Model [10], which describes the perception of the quality of the system from a user's perspective.

During the period of the evaluation of the different DataVaults platform versions, the consortium worked towards enriching the questionnaires available it is using to also encapsulate assessment questions relevant to ethics. As such, more questions have been posed to the engaged stakeholders. Moreover, as identified in the previous versions of the technical acceptance evaluation, during the final evaluation round the questionnaires were distributed to three target groups (e.g. demonstrator partners, data owners and data seekers), asking different set of questions per group, as this was not the case in the previous round due to the exercise being performed, as planned, in closed groups, consisting of people that had hands-on experience with the beta and alpha platform versions.

The next table presents the mean values of the final evaluation round for all questions asked to the different target groups (N/A cells mean that these questions were not asked to the specific target group)

Assessment Question	Data Owner	Data Seekers	Demonstrator Partner	Mean Value
General Questions				
Do the features offered by DataVaults **cover all tasks/objectives promised for this release**?	YES	YES	YES	**YES**
Do the function of DataVaults **accomplish the promised tasks** foreseen in this release?	YES	YES	YES	**YES**
Does the platform provide **accurate results** when it comes to its operations?	YES	YES	YES	**YES**
Can you **accurately reach your goals** with the system?	4,2	4	4,4	**4,2**
Do you think DataVaults **covers the intended purpose**?	4,4	4,2	4,7	**4,4**
Do you find DataVaults **useful**?	4,7	4,7	5	**4,8**
Do you **trust** DataVaults and its results?	4,2	5	5	**4,7**
Does DataVaults have **attributes that facilitate usability**?	Partially	Partially	YES	**Partially**
Do you believe DataVaults can be used by **disabled users**?	2,5	2,5	3,4	**2,8**
Do you believe **DataVaults can be used the system for other purposes** than the intended use?	N/A	N/A	YES	**YES**
PersonalApp Relevant Questions				

[10] https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en

Do the **aesthetics of the DataVaults PersonalApp** UI satisfy the needs of the user?	YES	N/A	YES	**YES**
Does **DataVaults PersonalApp please you** when you use it?	4,3	N/A	4,5	**4,4**
Do you feel that the DataVaults **PersonalApp provides a comfortable UI and workflow**?	4,2	N/A	4,7	**4,5**
To what extent does the DataVaults PersonalApp UI **allow you to exercise real control over your personal information**?	4,5	N/A	5	**4,8**
How much comfortable do you feel when using DataVaults as regards the **exercise of your rights** (information, access, rectification of inaccurate data, erasure, object, etc.)	4,6	N/A	5	**4,8**
Do you think that DataVaults PersonalApp is **easy-to-use for adjusting your privacy preferences**?	YES	N/A	YES	**YES**
Is DataVaults providing **clear notifications regarding your privacy risk exposure**?	YES	N/A	YES	**YES**
How much do you think that **DataVaults consent mechanism is fatiguing**? [11]	2,9	N/A	2,3	**2,6**
How much do you feel that **DataVaults is privacy and data protection friendly**?	4,5	N/A	4,8	**4,7**
Cloud based Platform relevant Questions				
Do the **aesthetics of the DataVaults Cloud Platform UI** satisfy the needs of the user?	N/A	YES	YES	**YES**
Do you feel that the **DataVaults Cloud Platform provides a comfortable UI and workflow**?	N/A	4,1	4,4	**4,25**
Does **DataVaults Cloud Platform please you** when you use it?	N/A	4	4,7	**4,35**

Table 29: Mean values of the final evaluation round

As it is drawn out of the questionnaires, the overall DataVaults experience is considered quite positive for all stakeholders engaged. In the next table a view of how these scores evolved during the project, showing a clear improvement over the course of the project.

Assessment Question	D6.3	D6.4	D6.5
General Questions			
Do the features offered by DataVaults **cover all tasks/objectives promised for this release**?	**YES**	**Partially**	**YES**
Do the function of DataVaults **accomplish the promised tasks** foreseen in this release?	**YES**	**YES**	**YES**
Does the platform provide **accurate results** when it comes to its operations?	**YES**	**YES**	**YES**
Can you **accurately reach your goals** with the system?	**4**	**3.6**	**4,2**
Do you think DataVaults **covers the intended purpose**?	**4**	**3.4**	**4,4**
Do you find DataVaults **useful**?	**4,3**	**3.6**	**4,8**
Do you **trust** DataVaults and its results?	**4,8**	**4**	**4,7**
Does DataVaults have **attributes that facilitate usability**?	**Partially**	**Partially**	**Partially**
Do you believe DataVaults can be used by **disabled users**?	**N/A**	**3**	**2,8**

[11] Lower score is better

Do you believe **DataVaults can be used the system for other purposes** than the intended use?	**NO**	**NO**	**YES**
PersonalApp Relevant Questions			
Do the **aesthetics of the DataVaults PersonalApp** UI satisfy the needs of the user?	**YES**	**YES**	**YES**
Does **DataVaults PersonalApp please you** when you use it?	**3,8**	**3,8**	**4,4**
Do you feel that the DataVaults **PersonalApp provides a comfortable UI and workflow**?	**4**	**4**	**4,5**
To what extent does the DataVaults PersonalApp UI **allow you to exercise real control over your personal information**?	**N/A**	**4,4**	**4,8**
How much comfortable do you feel when using DataVaults as regards the **exercise of your rights** (information, access, rectification of inaccurate data, erasure, object, etc.)	**N/A**	**4,4**	**4,8**
Do you think that DataVaults PersonalApp is **easy-to-use for adjusting your privacy preferences**?	**N/A**	**Partially**	**YES**
Is DataVaults providing **clear notifications regarding your privacy risk exposure**?	**N/A**	**N/A**	**YES**
How much do you think that DataVaults consent mechanism is fatiguing?	**N/A**	**3**	**2,6**
How much do you feel that **DataVaults is privacy and data protection friendly**?	**N/A**	**4,6**	**4,7**
Cloud based Platform relevant Questions			
Do the **aesthetics of the DataVaults Cloud Platform UI** satisfy the needs of the user?	**YES**	**YES**	**YES**
Do you feel that the **DataVaults Cloud Platform provides a comfortable UI and workflow**?	**4**	**4**	**4,3**
Does **DataVaults Cloud Platform please you** when you use it?	**3,8**	**4**	**4,4**

Table 30: Qualitative Evaluation Results during the three evaluation stages

One thing that the reader should consider is that the evaluation as performed in D6.3 was based on the alpha version of the platform, where the back-end functionality was limited, thus it is logical that in some cases the scores in D6.4 were lower, as various bugs and glitches came into the picture with the first release of the working prototype. However, numbers in D6.4 improved and reached or surpassed the initial values, pointing that the final product is according to the liking of stakeholders and is a product of high quality. This is visible in the following figures where the evolution of technical acceptance is provided for the 3 different versions of the demonstrators' operation.

Regarding the general evaluation questions for the technology acceptance of the overall solution, which were asked to all three groups, there is a noticeable decrease in various categories in D6.4, which have been attributed to the un-stable nature of the beta version of the platform. However, in the final evaluation (where the 0.50 platform version was evaluated, in all aspects the evaluation is better than the one initially performed with the alpha version (that measured more anticipated operations rather than real-life functionalities, as the base of the evaluation was semi-functional mock-ups). As such, we see a high score aspects that have to do with the realisation of the anticipated features, while the usefulness and the trust is also very high. Lower scores are provided to usability and to accessibility, however this fact is attributed to the fact that during the final evaluation, most responses came from data owners, which did not have the same traction with the system

such as the data seekers and the demonstrator organisation, that provided higher scores relevant to these dimensions.

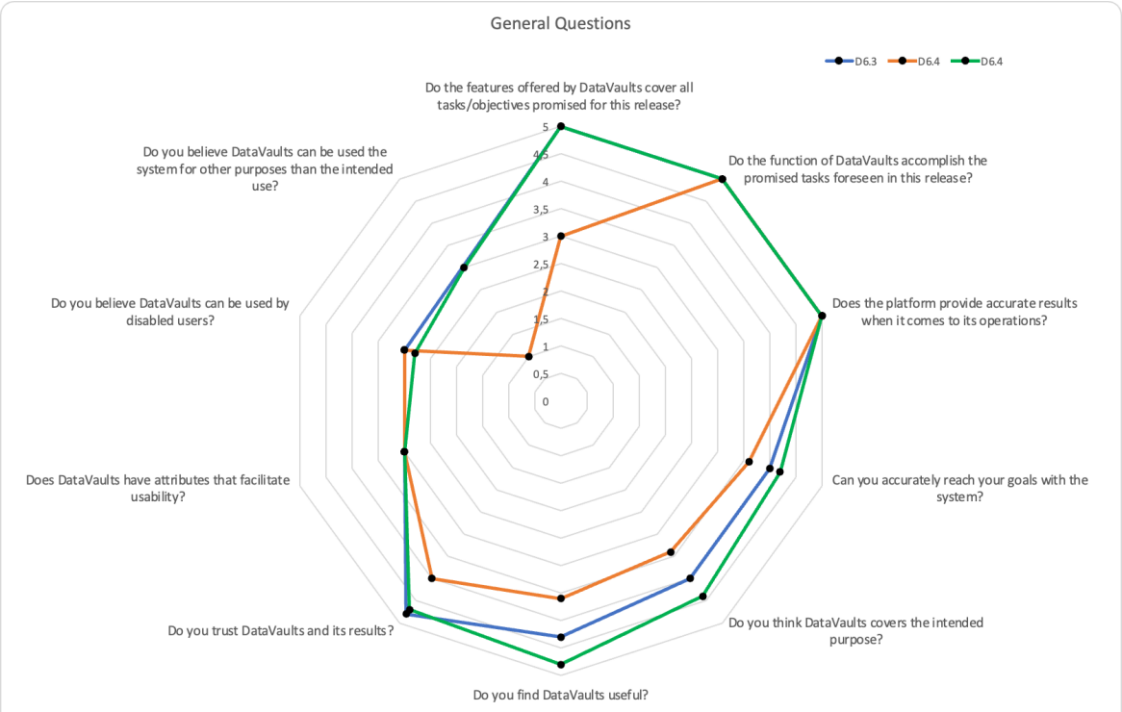

Figure 77: Evolution of TAM Questionnaire scores (General Questions)

Regarding the PersonalApp, responses came from Data Owners, as well as from Demonstration partners, who were involved in the testing and internal use of the systems. The questionnaire for this part of the platform tackled only a few aspects during the alpha evaluation version (as most questions required hands-on experience with the platform, which was not planned for the alpha version). As seen in the spider diagram below, in the final evaluation round the platform provided a better experience in all aspects analysed.

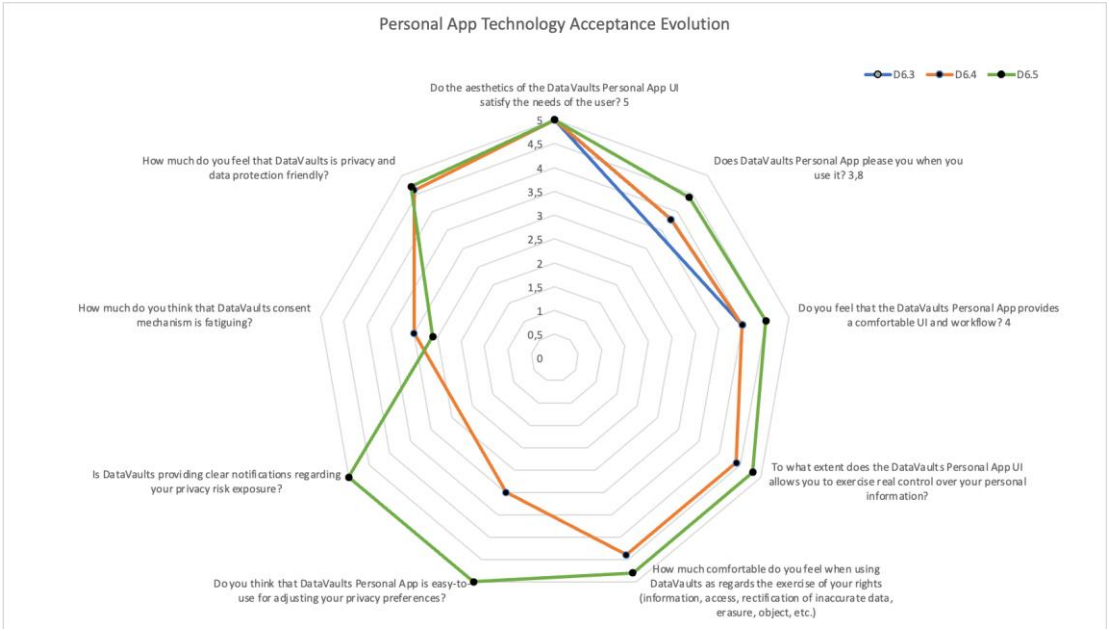

Figure 78: Evolution of TAM Questionnaire scores (PersonalApp Questions)

Finally, regarding the cloud platform, options about the ease-of-use and the UI have been positive, all ratings were improved as the final version included a refactored UI.

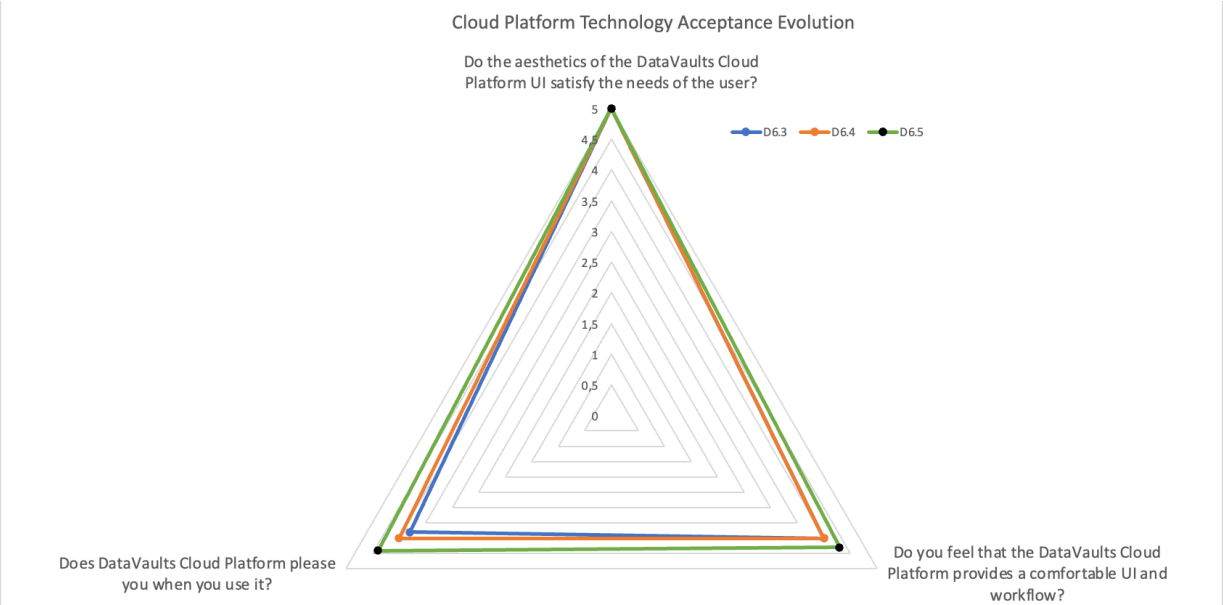

Figure 79: Evolution of TAM Questionnaire scores (General Questions) Cloud Platform

Summarising the overall technical evaluation from the perspective of the demonstrators, a continuous improvement relevant to the acceptance of the DataVaults platform and its usability has been recorded by the audience that used the platform, which suggests that the continuous integration activities of the project and the agile development format did meet the expectations set for delivering, in waves, a product that is continuously improved.

As the responses revealed, the scores coming from Data Owners were lower than those of Data Seekers and Demonstrators, however not at a magnitude that would raise concerns. This result was however expected, as Data Owners were offered with much more features that Data Seekers and did not have the traction with the platform that Demonstrator partners had. However, even lower, their scores were close to those of the other groups and given that their responses were much more in numbers compared to those of the other groups, allows us to assume that it was representative of the audience reached, thus having their scores so close to people that worked more with the platform, is something encouraging.

6 EVALUATIVE THINKING AND VALIDITY OF THE DATAVAULTS THEORY OF CHANGE

6.1 THE "EVALUATIVE THINKING" PROCESS

Underlying the evaluation process for DataVaults throughout the project was the concept of evaluative thinking.
An evaluation that reflects evaluative thinking is the systematic process of telling the DataVaults "story" by:

- Identifying assumptions about why we think the project will work and be a success.
- Determining what change we expect to see during and after we implement what we have set out to do in the Description of Action (DoA).
- Collecting and analysing data to help us understand what happened during the project.
- Communicating, interpreting, and reflecting on the results.
- Using these results and lessons learned to help make informed decisions to be able to plan for a successful exploitation after the project finishes.
- And one which can take on board influences from a wide range of stakeholders.

Part of the "storytelling" included the creation of a "Theory of Change" and a "Logic Model" which were set out in the Evaluation Plan in D6.1. The diagram capturing the DataVaults "Theory of Change" can be found in Appendix 1 of this document.

A general perception which was followed, was that evaluation should be designed into a project from the beginning. The DataVaults Evaluation Framework established that evaluation should be viewed as **a collaborative process that involves all of the stakeholders in various roles**, whilst it helped tell the story of the DataVaults project through a continuous cycle of asking, planning, and acting, reflecting and improving. We continually strove to make sure that findings were practical and useful for end users and able to inform decision-making and capacity building for further exploitation and sustainability. Indeed, evaluation can be regarded as a means of communication within the project.

In short, "*Evaluation is an objective process of understanding how a project or other intervention was implemented, what effects it had, for whom, how and why*".[12]

Thus, here we need to reference the most significant influences which we have taken on board in the work we have carried out, which may have re-shaped our view of the DataVaults "Theory of Change" as elaborated in Chapter 2 of D6.1.

In this section, we refer to the book "Personal Data-Smart Cities. How Cities can utilise their citizen's personal data to help them become climate neutral"[13] throughout. Although useful

[12] HM TREASURY, The Magenta Book, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf.
[13] https://smart-cities-marketplace.ec.europa.eu/media/2870

for dissemination purposes, it is a foundation for one pathway for exploitation and achieving critical mass. This is because the book reflects focussed research carried out by many DataVaults personnel addressing topics identified by stakeholders, including the reviewers, as being potential obstacles to deployment at scale for the DataVaults platform as a whole, within the context of a wider sharing of personal data across Europe, as envisaged by the Commission, the main stakeholder.

It is not a loose collection of essays, but a specifically commissioned and collected set of considered responses to issues identified by us and by some sixteen plus other projects with whom we built a relationship, focussed around the DataVaults-led "Citizen's Control of Personal Data" initiative within the EC's Smart Cities MarketPlace. It also embraced changes in policy focus over the period and hence the added emphasis on utilising personal data in the drive to become climate neutral.

6.2 Evaluative Thinking and Business Model Progress

The Evaluation Framework set out how we were to tackle how we have contributed to providing a **novel business model** and **contributions to the data economy**. But this was an aspect of the evaluation plan which could not be tackled in any detail until the final stages of the project, and as the technology matured. The objectives which were valid during the alpha phase and beta phases have been enhanced once we were able to eventually move from testing within a controlled closed group to engaging with a wider audience.

However, there have been three influences in particular, which have both confirmed our initial perceptions for the DataVaults business plan and its contribution to the data economy and provided us with added focus, which have led to broadening our outlook, embracing additional objectives and giving rise to further exploitation opportunities.

- One has been the feedback from reviewers and other stakeholders, including Commission officers.
- The second is linked to the valuable input gained through the DataVaults efforts in establishing and leading the Smart Cities MarketPlace "Citizen Control of Personal Data" Initiative. [14]
- And with DataVaults continuing close liaison with BDVA being the third.

In the final phases of the project, we took into account the suggestions made at the first review, both in the evolution of the DataVaults exploitation strategy and within the work led by DataVaults in the Smart Cities MarketPlace Initiative.

6.2.1 Smart Cities MarketPlace

At an early stage of the project, it was recognised that many of the partners had referred to smart cities in their individual exploitation strategies at the time the proposal was made and consolidating and adopting a joined-up approach would be beneficial. Individual partners talked of making products and offering services to smart cities, with two of the demonstration sites striving to be smart cities and with all the demonstrators being able to contribute to this goal in some way. The work started by DataVaults in the form of the Smart Cities MarketPlace

[14] https://smart-cities-marketplace.ec.europa.eu/action-clusters-and-initiatives

initiative helped to shape the project through providing new engaged stakeholders, each having conditions they would wish to see being met, whilst at the same time, providing answers to points raised by the existing stakeholders, including the reviewers. New stakeholders continue to be recruited, with new dissemination and exploitation channels emerging. The culmination of this work up to now (it will continue) has been the production of the book referred to above. The purpose of the "Citizen Centric approach to data"[15] initiative within the Marketplace can be summed up as **helping to build the conditions and relationships whereby the citizen will be willing to share personal data with a city and with other actors in the data economy.**

Whilst exploitation plans at the early stage of the project focussed on how personal data could be captured and utilised in the right circumstances, focus on the book concentrated more upon how personal data could enhance existing uses of data in deploying urban data platforms and in tackling the goal of becoming climate neutral.

Removing obstacles would help to grow the extremely valuable "personal data lake" which would then increase activity in the data economy and enrich existing data eco-systems. Success here would bring about tremendous benefits to the Data Economy as whole, and in particular to the digital SMEs within those cities, with access to data being regarded as more of a problem than access to finance. Growing an accessible "Personal Data Lake" will provide a means for improving most aspects of how data is currently utilised. And enable us to move more speedily to having a distinct European Data Model, of more benefit to wider society, rather than focussed on global commerce.

The topics we focussed upon included:

- Identifying cities and "sister projects" as exemplars in their general usage of data platforms and to highlight what value there is in adding personal data to that being utilised.
- Having a citizen-centric focus by creating impactful use cases and easy to use services/Apps facilitating citizens to remain in control of their data, which is at the core of DataVaults.
- Capture/pilot collaborative/joint business case; developing methods and tools that will help multiple cities adopt.[16]
- Providing answers to "legal and ethical" questions arising.

The results of these collective deliberations became the seventeen chapters of the book which was produced.

[15] https://smart-cities-marketplace.ec.europa.eu/action-clusters-and-initiatives/action-clusters/citizen-focus/citizens-control-personal-data

[16] " https://smart-cities-marketplace.ec.europa.eu/sites/default/files/2021-03/Notes%20-%20Workshop%20Citizen%20Control%20workshop%20on%2025.02.2021_0.pdf

6.2.2 Reviewers as stakeholders: Related comments from the first review

6.2.2.1 *Smart city applications*

"While some transport-related uses of the user location data are mentioned in some of the use cases/demos, it is perceived that the potential of use cases that such data can support in smart city applications is currently underexploited (the exploitation plans are not covering all the possibilities in this area)"

Whilst not being presented in detail at an earlier stage of the project, work has been ongoing in interacting with the stakeholders operating in the smart cities' environment. Efforts to turn those interested in the use of their citizen's personal data within a smart city context have been engaged with. Just as the chapters in the book produced by the consortium were purposefully commissioned, so were the cities and projects which were invited to share their knowledge with us. The network of these cities will become a valuable tool after the project has finished, and as with several of the points covered in this section, is covered in more detail in D7.4 Final Project Exploitation Report. Comments have re-enforced our position and the strategy adopted, with mobility being seen as a key component in the future DataVaults offering within the smart cities' context.

Work has progressed within the smart city demonstrators of Piraeus/Olympiakos and Prato in creating exemplars for the utilisation of adding the value of location data to enhance transport planning and the insistence on having a "DataVaults" enabled questionnaire process, to feed additional rich data into the planning process, was indicative of the recognition of building on this aspect of DataVaults.

The approach taken in chapter 6 of the DataVaults produced book draws attention as to how accessing additional personal data can only enhance current data usage within a smart city environment.

6.2.2.2 *"Marketing focus on societal challenges"*

The focus on societal challenges within a smart city environment has been boosted by the changing emphasis on achieving the target set by many cities of becoming climate neutral by 2030, providing a focus and the ability to demonstrate that by adding personal data to those existing data sources which smart cities are already utilising, greater benefit is derived. The launch of the book and resulting dissemination activities have focused on the issue of adding personal data to existing data applications in the drive towards attaining net-zero, be it in energy, mobility or generally increasing citizen participation in the process.

6.2.2.3 *Market-economy definition of the value of data is potentially problematic.*

We have organised activity within the Smart Cities MarketPlace initiative to cover this topic resulting in a chapter in the DataVaults book, commissioned from the SAFEDEED project, tasked with working on this problem. Work will continue along the lines of creating new models utilising the "Story of Data" to establish novel ways of valuing data, given the market failure and the nature of data being re-usable. The "What Next" chapters set out how a group of cities will collaborate to tackle such issues highlighted by stakeholders. It is in the same context that the concept of a citizen sharing their data, which they willingly give to the large

non-European corporations, with their city authorities, to aid a common cause can be seen. This aspect is at the core of the work with smart cities but still maintaining the ability for a citizen to also gain reward in kind- free transport for example, or by helping the local data economy grow for real rewards by selling data, given that SMEs find it easier to raise finance than they do to access data. Hopefully removing the ironic situation of a DataVaults city having to purchase back data derived from its citizens.

6.2.2.4 *New business models emerging from such data owner sharing arrangements.*

Extensive work has taken place with sister projects such as DUET[17], RUGGEDISED[18] and SAFEDEED[19] and with the EC Joint Research Centre, within the framework of the MarketPlace initiative to provide answers regarding how a DataVaults platform might be deployed, what staffing arrangements might be required, how the platform should be governed and again covering the value of data and new forms of business models, particularly in relation to cities coming together to collectively exploit personal data on a larger scale.

6.2.2.5 *"Reducing the user burden when dealing with consent».*

DataVaults has engaged with initiatives looking to standardise how citizens should interact with the consent process. However, these are in their development stage, but we will continue to be in contact with them as part of the continuing Smart Cities MarketPlace initiative.

6.2.2.6 *"DataVaults can potentially disrupt the monopolies that are on the rise."*

Attention was drawn to the fact that the business models of the huge data corporations might be in direct conflict with the principles, intentions, and business models of DataVaults. Whilst this is further discussed in D7.4, a driving force behind the MarketPlace initiative has been to create an understanding that new sources of personal data are available, enhancing existing use of data supporting all aspects of a smart city, whilst also strengthening the local data economy is a potential alternative solution. Data from citizens at scale across the cities and the follower cities aiming to be climate neutral by 2030 is a route to achieving scale, independent of the large corporations and their processes, accompanied by the legislative changes in progress. By addressing "Net-zero cities", a pathway to achieving critical mass opens.

6.2.2.7 *"The role, value, and involvement of so-called 2nd tier economic operators are weakly defined."*

As reported in Section 3 and in deliverable D7.4, effort has been directed at ascertaining the optimum role for this group of stakeholders, which are entities that are not interested to onboard DataVaults, but to acquire services that can be produced by a data seeker who can collect and analyse personal data and generate relevant reports.

By different activities through dissemination and invitation of collaborators to the demonstrators, external to the consortium entities had been engaged into discussions

[17] Home | digital urban european twins (digitalurbantwins.com)
[18] RUGGEDISED - Smart city lighthouse project | LEGACY
[19] SAFE-DEED | Safe Deed

relevant to the offerings of the platform and as it turned out, some of them are not interested to be directly engaged as stakeholders, but as recipients of services/intelligence that a data seeker or platform operator can provide to them. As an example, we mention here the sponsors which were approached by the Olympiacos demonstrator that would be interested to consume directly intelligence provided by Olympiacos (as a stakeholder), or the company approached by MIWenergia which declared its interest to consume reports and analytics.

6.2.3 Commission's Changing emphasis regarding smart cities

The change of emphasis from the Commission referred to above, regarding the direction smart cities should be taking in future has been acted upon. The new initiative focussed on having a core group of 100 climate neutral cities by 2030 which will act as lighthouse for 350 more cities. Funding will be sought to enable a group of these cities, (Prato being one of the 100 cities), to enable them to investigate in depth many of the topics above which were raised in the first review and contributing to what best practice in dealing with citizen's personal data should be.

Further, there has been a "stock-take" of the plethora of networks and collaborative mechanisms which have emerged over the years, so as to have more effective intervention, given the short time scale set for getting to climate –neutral status. DataVaults is positioning itself within these anticipated changes.

Primarily interaction with the European Commission, in addition to the normal reporting mechanisms, was with the Smart Cities MarketPlace which was a joint initiative created unusually by three DGs, DG ENER, DG Move and DG CNECT. Whilst the book written in association with this initiative was strongly supported and encouraged by DG ENERG, it was launched by personnel from DG CNECT.

A focus is also being provided by the ongoing deliberations by DG DIGIT surrounding proposed data space covering personal data and the development of a "Smart and Sustainable Cities DataSpace." DataVaults has been invited to attend the second Cluster Meeting which is helping to determine what shape this will take. Alongside this is a close relationship with those working towards the "Personal Data Dataspace".

More indirect was the feedback into policy-making processes through BDVA.

6.2.3.1 *Interoperability*

And as a final example of how the DataVaults "Theory of Change" has been utilised to prompts from stakeholders, work is planned in the final phase to act on advice from Commission officials to engage with the interoperability work coming out of Livingin.eu[20] and with the MyData[21] movement, which was acted upon.

In pursuit of this, work has commenced with Open and Agile Smart Cities and Livingin.EU to experiment with overcoming potential interoperability problems which might slow down the advance. Plans have been formulated to deploy the Minimum Interoperability Mechanisms

[20] "https://living-in.eu/"
[21] "https://mydata.org/about/"

currently being developed within Living-in-EU but trials with the DataPorts project and MyData in Piraeus have had to be postponed awaiting the final delivery of MIMS4 which is still being finalised.

The Smart Cities Marketplace has entered a new term which will work with a different, more agile and flatter structure emerging from its restructuring. Many initiatives will be discontinued. A clear output-driven approach is in place, to support the Explore-Shape-Deal funnel. This is of advantage to DataVaults as a route to wider take-up, as the initiative it established is being encouraged to remain active and this advice below is being heeded.

“Citizens’ control of Personal Data”: it is proposed to continue this initiative under the condition that possible collaboration with the Smart Sustainable Cities Data Space and the Living-in.EU initiative of DG-CNECT is explored.”

“The Living-in.EU Minimal Interoperability Mechanism Plus includes an interoperability mechanism that is focused on personal data management. Collaboration with the CSA on the smart communities’ data space and the Living-in.EU CSA is strongly encouraged to ensure technical compatibility.”

Prato, in its development of their digital wallet, have ensured that DataVaults looks also in line with the EIDAS digital wallet that is under development.

6.2.4 Policy making at the global level.

In addition to efforts to contribute to EU policy making, DataVaults has contributed to the consultations held by the United Nations on its future digital strategy- the Global Digital Compact[22] alongside those from the EU and Google. It has also held discussions with UN Habitat around the work they plan to carry out around utilising whilst protecting personal data.

[22] Global Digital Compact - submissions | Office of the Secretary-General's Envoy on Technology (un.org)

7 CONCLUSIONS

The overall view is that progress continued to be made in achieving the overall DataVaults objectives and that by the time the technology had matured, the demonstration sites were well-prepared for the important final phase of the project. The demonstration sites carried out their roles as described in the planning documents and in the DoA and the results of the successful trials were highlighted in Chapter 3.

Pleasing were the variety of comments made by the demonstration sites on how useful DataVaults will be to them in the future. Andaman highlighted how fit for purpose it was in relation to conducting medical trials as well as enhancing their existing personal health application.

Piraeus concluded that it had real value for the municipality stating: “The ease and speed in collecting information with respect to the current ways of gathering data is unmatched. Moreover, the ability to obtain personal information in a GDPR compliant way, without tiresome procedures and worries about the legal aspects of the data acquiring process is a strong reason for using the platform.”

Similarly, the ability to describe the exact type of required information and its accuracy was noted in comparison with existing mechanisms. Piraeus also gained from the successful trialling by Olympiacos in working on data that revolve around the behaviour and the profile of its clients (e.g., the fans/members). For example, deriving value from knowledge relevant to the preference of fans regarding mobility and commuting options (for example the case of municipal and traffic management authorities, or of local businesses), and regarding their spending power and behaviour on match days. For Olympiacos, it will also allow the club to open new revenue streams.

Regarding mobility planning in Prato: “Currently, the administration gets such information by means of specific and expensive sensors, often put in place for a limited period of time, while the DataVaults approach will enable getting data in a more continuous and rather simple way, even by profiling users and reducing costs for the administration in planning new solutions for the city mobility.”

Similarly, the tool produced almost as a by-product for DataVaults, of having an enhanced form of questionnaire available, proved valuable across the board, enabling the gathering of much richer personal feedback from citizens to be gathered. As Prato pointed out in their case: “Currently, this activity is carried out by external companies that are usually rather expensive and this is the reason why such investigations are not very often carried out, while the DataVaults approach can be very useful for the setup of more continuous surveys and monitoring of results.”

MIWenergia have demonstrated that there is a market for the data produced and for analysed data and sees the benefit from the use of the platform via the collection of personal and energy data that can help the company to design personalized services, special targeted offers, and for the profiling of clients and potential clients.

Whilst the technical evaluation was primarily reported in D5.6, Chapter 5 concluded that the demonstrators were able to continue the progress made throughout the testing of DataVaults throughout the project. Previous deliverables had reported how progress was made between the alpha and beta versions and this deliverable reports on this continuing trend and further improvements to the technology and its usage.

Having moved on from where the technology had limitations preventing a full evaluation of the full system, the progress made in the final period enabled full testing to occur. The demonstrators were able to utilise the fully rolled-out blockchain infrastructure, and evaluate all aspects of the system, where a more thorough evaluation of the UX was performed, as well as of the facilities to be used for providing information about sharing transactions and for distributing value amongst stakeholders.

Whilst the project has visibly created the technology and demonstrated it in high-profile applications across the EU, the work led by ETA covering all the security, legal and aspects preventing the citizens from willingly sharing their personal data has been ongoing. From Chapter 2, we can see that the necessary technical, ethical, and legal steps have been put in place to ensure that we can confidently proceed with deploying the final version of the platform at the demonstration sites.

Successful stakeholder interactions have provided “food for thought” in progressing the “evaluative thinking” and utilising the “Theory of Change” underpinning the DataVaults project, confirming that we are on track. They have further paved the way to assist in exploitation.

The demonstration sites themselves are well-prepared for the next phase of the project’s exploitation, while progress in their scenarios has been performed in terms of building connectors and progressing their envisaged test cases, mostly by engaging internal stakeholders and preparing the ground for opening DataVaults to the public.

Lessons were learned throughout the demonstration process and have been reported, with the most valuable lessons learned being reported elsewhere in D6.6 DataVaults Scaleup Roadmap and Key Takeaways Documentation and lessons learnt from the DataVaults project, constituting methodological adoption guidelines for the utilisation of the platform.

It was concluded that the methods and functions offered do generate a very high feeling of trust, and in general the features offered are in line with what has been expected by the demonstrators based on the development backlog. All the recommendations made in D9.2, the requirements elicited in D2.1 and augmented in D2.3 have been properly followed by the technical team of the project during the design and development of the DataVaults technologies and by the demonstrators during the validation phase.

The DataVaults demonstrators’ processes and operations adequately implemented the ethical procedures outlined in D9.2 and in WP10 deliverables. Furthermore, the demonstrators strictly adhered to the Ethical and Legal requirements specifically set for them in the mentioned deliverables, such as the need to follow adequate consent procedures and recruiting procedures and to use suitable tools and safeguards.

These considerations, and in particular the outcomes of the EDPIAs and the fulfilment of the requirements regarding the DataVaults technology elicited in D2.2 and D2.3, confirm that the DataVaults platform, its app and the other technological assets generated in this research are legally compliant and ethically sound and give rise to a trusted, secure privacy-friendly and citizen-respectful data sharing environment.

It can be concluded that the DataVaults partners succeeded in achieving what they set out to do and have paved the way for this work to be built upon and exploited.

APPENDIX 1: EXHAUSTIVE LIST OF POTENTIAL QUESTIONS RAISED THROUGH THE LOGIC MODELS, THE DOA AND PROJECT DELIVERABLES INITIALLY CREATED IN D6.1.

The intention in producing these questions, which were derived from the DoA and the relevant "call notice", was not to neatly measure progress, but to give a guiding overview and to maintain focus on what we were doing and what should be achieved, dovetailing with the project's Theory of Change. They collectively asked whether we were doing as we promised in taking the direction which the project took.

Questions which can be asked of the DataVaults project to refine for evaluation relating to legal and ethical issues.	**Response**
Chapter 2 and the deliverables D2.1 and D2.3, D9.2 and the WP10 deliverables have covered these topics and have underpinned the security, privacy, legal and ethical aspects of the project.	
"Personal data platforms shall ensure respect of prevailing legislation and allow data subjects and data owners to remain in control of their data and its subsequent use."	Yes
"Additional efforts in the future could be useful to further improve the user-friendliness of the User Interface in order to enhance the effectiveness of the individual's control over their data."	
"…Develop privacy metrics that are easy to understand for data subjects and contribute to the economic value of data by allowing privacy-preserving integration of independently developed data sources."	Yes
"… Conditions of use and practical arrangements of data sharing should be regulated."	Yes
Did we successfully link novel trusted and security-by-design data mining, management, analysis and sharing techniques, with legislation- and ethics-driven functions?	Yes
Did this facilitate both privacy and trust preservation?	Yes
Did this facilitate risk situational awareness?	Yes
Has there been an Improvement in Trust?	Yes
"This is the overall perception from the final evaluation with users/data owners, though it ran only for a limited time as part of the project, therefore additional investigation of this aspect would be opportune in the future."	
Has a mechanism for taking back citizen control been provided and tested?	Yes
Are individuals receiving a fair share of their data value? (*How have we determined what is fair?)*	Yes
Are individuals receiving a fair share of their data value for second level use? (*How have we determined what is fair?)*	Yes comparatively.
"The technology has been conceived and designed taking into account fairness and human empowerment. However, the validation with users/data owners ran only for a limited time within a project, therefore additional investigation of this aspect would be opportune in the future." The market mechanism alone is not the answer.	

Has there been a contribution to the Basis for Privacy, Ethics and IPR?	Yes
Is there compliance with all relevant regulations?	Yes
Do we contribute to and are we aligned with:	
Ethic driven standards	Yes
Projects and regulatory tools	Yes
Fundamental Rights and Well-being	Yes
Is the PersonalApp compliant with EU regulations and national laws?	Yes
Has the PersonalApp been successfully used by individuals for storing, collecting and sharing data and what was their experience?	Yes
"The technology has been conceived, designed and developed taking into account the legal and ethical requirements. However, the validation with users/data owners ran only for a limited time, therefore additional investigation of this aspect would be opportune in the future"	
Was the personal app consent mechanism clear and well-received?	Yes
Are there fair compensation models for all the actors of the value chain available?	Yes
"Though this aspect needs to be further investigated and tackled in the future, taking into account the context-specific factors."	
Did we provide an Ethics monitoring framework?	Yes
Have we given improved control and awareness of how a citizen's data is shared and managed?	Yes
Have we provided an acceptable "Remuneration Scheme" based on the data produced and shared?	Yes
Improving Privacy Risk Exposure Awareness for Individuals when sharing Personal Data *already established*	Yes, though this aspect was not extensively evaluated
Improvement of Individuals Knowledge on Personal Data Safeguarding	Yes
Did we address and improve the situation where there may be reluctance to personal data sharing services via DataVaults?	Yes
Increase of the value of personal data attributed back to owners	Yes
Have we demonstrated that Citizens' trust is improved as privacy-aware transparency and control features are increasingly streamlined across data platforms and Big Data applications?	Partially
The technology has been conceived, designed and developed taking into account the legal and ethical requirements, including those regarding privacy-aware transparency and control features. However, the validation with users/data owners ran only for a limited time, therefore additional investigation of this aspect would be opportune in the future	
Have we demonstrated that we have Improved Privacy Risk Exposure Awareness for Individuals when sharing Personal Data?	As above
Can we show an Improvement of Individual's Knowledge on Personal Data Safeguarding?	Yes
Have we overcome reluctance to personal data sharing services via DataVaults?	Yes, partially

Questions which can be asked of the DataVaults project to refine for evaluation relating to business model issues.	**Response**
Chapter 5 and the deliverables in WP7 have covered many of these topics	
Has a secure trusted platform been provided and tested?	Yes
Is there evidence that GDPR related costs in investment decisions have been reduced for industry?	Yes
Business Innovation	
Cross-domain collaboration	Yes
Has Increased scale been demonstrated from small existing initiatives?	Yes
Have entrepreneurs been able to develop their own goals and operations?	No
Have we created a novel business model for personal data and insights sharing where data is valued based on different modalities and is attributed rightful owners:	Yes
Has an inclusive analysis regarding DataVaults Trusted Data Management and Sharing Principles been carried out?	Yes
Has the DataVaults Personal Data Sharing Business Model been developed?	Yes
Has the DataVaults Value Distribution Method been delivered?	Yes
Have Smart Contract Patterns and Templates for Stakeholder Collaboration and SLAs been provided?	Yes
Has a PESTLE analysis of the DataVaults ecosystem been delivered?	Yes
Has a SWOT analysis of the DataVaults platform been provided?	Yes
Have 3 different Business Models been templated for personal data value sharing?	Yes
Has a distributed ledger platform to facilitate transactions, been provided?	Yes
Has there been a consortium wide exploitation plan delivered?	Yes
Has an inclusive financial strategy plan with cost breakdowns and future projections been delivered?	Yes
Has a consortium wide sustainability plan been delivered?	Yes
Are individual partner's business models and exploitation plans in place?	Yes
Have we cultivated a trusted sustainable and ever-growing ecosystem, with industry offers expanded by citizen-controlled access to more and varied data?	Yes
Did it propel the creation of a joint venture of personal data owners and data seeking organisations?	No. Explained in D7.4
Did we "link to and bring in industrial data providers, (not necessarily as consortium members), that will populate the platforms."	Partially. Links have been made only
Have we provided Data Industries with easier and seamless access to personal data?	Yes
Have we provided acceptably secure and privacy aware guarantees for this access?	Yes
Have we significantly increased opportunities related to integrated data and data integration services' provision?	Yes
Have we provided more evidence-based analytics to support their strategic business and operational decisions?	Yes

Have we provided innovative and more effective products and services?	Yes
Have we significantly reduced time to market for new products and services?	Yes
Have we significantly increased business opportunities related to innovative services and apps?	Yes
Have we provided easier and seamless access to constantly growing volumes of cross-sectorial multilingual big data?	Yes
Have we provided new business opportunities related to building on top of existing solutions?	Yes
Have we provided improved and fast access to personal data allowing data scientists to focus on experiments development, rather than investing time on data management and collection issues that need to be tackled due to regulation?	Yes
Have we provided a secure environment for experimentation with sensitive personal data structures?	Yes
Have we demonstrated that personal data protection is improved, and compliance with the General Data Protection Regulation (and other relevant legislation) is made easier for economic operators, including SMEs?	Yes
Have we reduced investments for personal data handling for enterprises using DataVaults to access and analyse personal data?	Yes
Have we demonstrated better value-creation from personal and proprietary/industrial data?	Yes
Have we demonstrated a 20% annual increase in the number of data provider organisations in the personal and industrial data platforms?	Long-term Impact Expected
Have we demonstrated a 50% annual increase in number of users (data subjects) in the personal data platforms?	Long-term Impact Expected
Have we demonstrated 20% annual increase in volume of business (turnover) channelled through the platforms?	Long-term Impact Expected
Have we shown a lowered effort to handle GDPR issues for SMEs?	Yes
Reduced investments for personal data handling for enterprises using DataVaults to access and analyse personal data	Yes
Improved access to personal data for economic operators	Yes
Increase of revenue for EU data companies	Yes
Enlarge the base of EU data scientists/engineers	Yes
Improved access to personal data for economic operators	Yes
Increase of value of reports and services based on personal	Yes
New services offerings per year created for economic operators	Yes
Annual increase in the number of data provider organisations in the personal and industrial data platforms	Yes
Annual increase in number of users (data subjects) in the personal data platforms	Yes
Annual.inc in volume of business (turnover) channelled through the platforms	Yes
Service quality and experience improvement through personalisation	Yes

Questions which can be asked of the DataVaults project to refine for evaluation relating to the product delivered.	**Response**
Chapter 5 and the deliverables in WP3,4,5 have covered many of these topics etc	
Have we made a good product?	Yes
Have the Data Sharing means provided been tested successfully?	Yes
Have the Data Analytics been successful?	Yes
How well have we combined fragmented and domain specific data?	Yes
Have we successfully demonstrated the use of data derived from the wide variety of sources available?	
Recent innovations in sensors	Yes
activity tracking through wearable devices	Yes
Internet of Things	Yes
Cyber- Physical Systems (CPS) technologies	No
wearables	Yes
data APIs	Yes
historical data	Yes
social network data	Yes
activity trackers	Yes
health records	Yes
demographic profiles	Yes
Others used in the demonstrations	Yes
Have smart contracts within the data chain been successfully tested?	Yes
Has the Cloud-based pan-European Personal Data Platform and infrastructure been tested successfully?	Yes
And has the PersonalApp been tested successfully?	Yes
Has it High Security and Privacy?	Yes
For personal data and derivatives Have each of the following components been successfully delivered?	
1 cloud-based Data Management and Analytics platform,	Yes
1 Personal Data Management and Analytics App,	Yes
1 Open-source library of the Personal Data App components	Yes
1 Business Validation and Impact Assessment Report	Yes
Have the Impact Metrics targets been achieved?	Partially. See D6.5 and D7.4
Have we delivered the technical solution comprising of:	
Secure and trusted Data Management and Analytics cloud based platform as a Service?	Yes
Personal Data Apps, for storing, managing, sharing and monetizing over personal data (derivatives) which can be used by any individual with the aim to capitalise on the real value of his personal data, without dropping control of ownership or loosing track of the usage methods, providing also constant awareness of the privacy, security and risks he may be exposed?	
Has the DataVaults cloud-based platform been successfully tested?	Yes
Has the DataVaults mobile PersonalApp been tested?	Yes
Are the DataVaults PersonalApp libraries of the standard desired?	Yes

Is the Data Model as desired?	Yes
Has the DataVaults Open API been tested?	Yes
Has the Data brokerage engine with two layers been tested?	Yes
Have we produced the Documentation and Usage manuals?	Yes
Have we produced an expandable knowledge representation (in the form of a knowledge graph or ontology) for the personal data?	Yes
Have we provided and tested a secure cloud-based storage facility?	Yes
Have we provided and tested a cloud-based data analytics engine?	Yes
Have we provided and tested an access control engine?	Yes
Have we provided and tested a risk management service?	Yes
Have we provided and tested a visualization library?	Yes
Have we provided and tested a personal data catalogue?	Yes
Have we provided and tested a business brokerage engine?	Yes
Have we integrated existing approaches, tools, libraries and components that allow: • Handling of personal data in the way they should be preserved, accessed, valued, and controllably shared? • Guaranteeing high quality results which can support rapid prototyping, traction generation, fast market entry and sustainability?	Yes
Did we provide?	
Modularised Services and Tools for data management and sharing as part of the platform,	Yes
a unified data management service to interconnect all other components	Yes
improvement and integration of technical data infrastructure solutions supporting both secure and trusted data exchange and retention,	Yes
a novel paradigm for the documentation and IPR handling of conducted exchanges	Yes
Did we provide the platform architecture?	Yes
Did we provide designs of all software bundles?	Yes
Did we integrate at least 8 TRL>7 technologies?	Yes
Did we offer support for cutting edge technologies for security and trust by design?	Yes
Did we offer support for modern analytics algorithms on plain, multiplexed and encrypted data?	Yes
Scientific and Innovations Objective I:	
Did we deliver an innovative, secure, privacy preserving, IPR respecting, and fair compensation data exchange methodology?	Yes
Did we define the Value Chain of Personal Datasets and Data sources?	Yes
Did we provide a semantic representation of Personal Data?	Yes
Did we update existing semantic vocabularies and make contributions to LOD?	Yes
Did we provide data analysis algorithms?	Yes
Did this give easy access to and usage of valuable information?	Yes
Did we provide a security and privacy by design Personal Data lifecycle Management framework?	Yes
Did we provide an Assets Brokerage methodology?	Yes

Did we provide methods to isolate data and make them searchable even when encrypted?	Yes
Did we provide methods to share data at different levels and modalities?	Yes
Did we provide methods to calculate risk exposure?	Yes
Did we provide a Data Access Framework?	Yes
Did we provide an Assets Brokerage Engine?	Yes
Did we provide a Risk Exposure Dashboard?	Yes
Did we provide a Secure Data Management Environment on cloud and on App?	Yes
Have we delivered the following services?	
Holistic personal data management services, including collection, mining processing, normalization, formatting and availability at individuals' personal devices level as well as on secure data vaults on the cloud	Yes
Smart interlinking of personal data to open, linked as well as proprietary data following Linked Data principles and openly (re-)publishing non-sensitive and business critical information to the LOD community	Yes
Novel data security and cryptography, data anonymisation and privacy preservation, remote attestation and trusted data exchange through the utilisation of TPM technologies between the Personal DataVaults and the DataVaults cloud-based engine	Yes
Privacy risk assessment methods that offer a "situational awareness" picture to individuals with easy to understand privacy metrics, revealing the true risk exposure factor of individuals based on the shared data	Yes
Privacy preserving and data security retention mechanisms, to accommodate the generation of anonymised "digital twins" of individuals, as well as specimen clusters ("persona groups") powering group analytics that contain valuable insights without violating privacy principles	Yes
A twin fold data brokerage engine to cater for IPR and data license safeguarding, documenting transactions in a privacy preserving, yet undisputable and unforgeable manner, facilitating compensations schemes with third parties (that support the shift to future monetisation streams) through the instantiation of multi-layer real-time micro-contracts specifically tailored to the needs of data sharing, redistribution and utilisation, constructing a bridge between personal data and industrial data platforms.	Yes
Smart balancing of analytics methods to accommodate Edge Analytics as well as centralised operations depending on the degree of data volume, velocity and variety, always in conjunction with the security and privacy modalities allowed by the individual for each kind of analysis	Partially
Provision of intuitive analytics, reports, smart dashboards and visualizations tailored to the needs of each stakeholder of the domain, including the individual, as well as generic ones for wider use by any interested organisation and by the public	Yes
Are all the required Data sources from WP1 covered?	Yes
Do we have more effective services?	Yes
Do we have more efficient services?	Yes
Do we have more value-adding services?	Yes
Numbers of second tier operators per demo?	1

Novel services?	Yes
The 4vs of big data: Volume, Variety, Veracity, Velocity?	Yes
Questions which can be asked of the DataVaults project to refine for evaluation relating to stakeholder interaction issues.	**Response**
Chapter 4 and the deliverables in WP8 have covered many of these topics supported by all the activity recorded in the dissemination tracker.	
Did we deliver:	
DataVaults Public Showcase and Web Presence,	Yes
Marketing Kit,	Yes
Exploitation and Marketing Plan,	Yes
Dissemination and Stakeholders' Engagement Plan,	Yes
Events and Workshops,	Yes
Publications and Press material	Yes
Collaboration with other projects and businesses	Yes
Achieved communication targets	Yes
Did we run 200 questionnaires for needs elicitation?	Yes
Did WP8 achieve its own KPIs and targets?	Yes
Questions which can be asked of the DataVaults project for evaluation relating to technical issues.	**Response**
Chapter 5 and the deliverables in WPs 4,3,5 have covered all these topics.	
Risk exposure metrics –user needs and evaluation	Yes
Searchable data catalogue	Yes
Methods of making data available	Yes
Edge Cloud	Yes
Amount to share	Yes
Encrypted	Yes
Anonymisation	Yes
User interface for distributed ledger contracts engine	Yes
Personal DataVaults modules	
Data Fetcher and Transformation mechanism	Yes
Data Schema Repository	Yes
Secure Storage facility	Yes
Policy Access Editor,	Yes
Privacy Metrics Dashboard	Yes
Data Anonymiser and Identities Wallet,	Yes
Data Publisher and the TPM DAA module.	Yes
Edge Analytics Engine	Yes
Data Request Service Resolver	Yes
Data Picker	Yes
Personal DataVaults Wallet,	Yes
Private Ledger	Yes
Cloud based DataVaults platform modules	
Data Fetcher and Transformation	Yes
Data Schema Repository	Yes
Replicas of the Personal Data Storage	Yes
Encrypted Searchable Data Lake	Yes

Access Policy Engine	Yes
Indexing Service	Yes
Risk Management Monitor,	Yes
Data Policies Enforcement Services	Yes
Anonymizer bundle, which includes the Digital Twin Generator and the Persona Group Generator.	Yes
Scalable Data Analytics Containers	Yes
Secure Analytics Playground	Yes
Visualisation Dashboard	Yes
Data Request Service	Yes
Open API	Yes
Query Builder and Data Explorer	Yes
DataStream and Contract Composer	Yes
Open Ledger	Yes
Private Ledger	Yes
Personal Secure Data Management Functions and Edge Analytics bundle	Yes
Trusted Cloud-based Secure Analytics and Data Retention Services bundle	Yes
Trusted Data Sharing and Contract Negation Features Bundle	Yes
Questions which can be asked of the DataVaults project for evaluation relating to the piloting activities.	**Response**
Chapter 3 and the previous deliverables in WP6 have covered all these topics and further lessons learned are being reported in D6.6.	
Pilot Key Performance Indicators	
DEMONSTRATOR #1 – SPORTS AND ACTIVITY PERSONAL DATA (OLYMPIACOS)	
Increase in Stakeholder Trust	98%
More effective management of members and fans data	94%
More effective management of sport activity data.	100%
Fans profiling	Complete Profile of Fans
Whereabouts of fans after events	3 main routes
Access to daily activity data of athletes	1 Ergometric online profile dataset per athlete
More registered members	Long-Term Impact KPI
More “active members”	Long-Term Impact KPI
More “active fans”	Long-Term Impact KPI
Increased sponsorship revenue	Long-Term Impact KPI
DEMONSTRATOR #2 – STRENTHENING ENTREPRENEURSHIP AND MOBILITY (PIRAEUS)	
To increase the number of citizens actively sharing data.	7271
New municipal services.	4
Improved citizen’s satisfaction with services.	26%

Touristic activity.	880.416
Improved local commercial activity.	Over 8.000.000€
Increase in the number of customers entering the local stores	23%
Increase in revenues of the local shops participating in the pilot	18%
Number of entrepreneurs involved	22
Number of shared datasets	340
Decrease in tie required to reach the sports venue	1%
Decrease in time to park around the sports venue	1%
Number of tourists and citizens participating	150
Number of data analysis procedures	2
Number of actions taken by the local Destination Management Organization based on DataVaults data	0
DEMONSTRATOR #3 – HEALTHCARE DATA RETENTION AND SHARING (ANDAMAN7)	
To increase active users operating Andaman7 *Scenarios: a and b*	Registered users: 38240 Average active users: 1,200
To increase the number of datatypes used by Andaman7. *Scenario: b*	Increased to 243 to mitigate
Increase of Volume of data per/ category *Scenarios: b*	Increase of 29% per user (all categories) to mitigate
Generation of new services offered through Andaman7. *Scenarios: a and b*	4 services added
DEMONSTRATOR #4 – SMARTHOME PERSONAL ENERGY DATA (MIWENERGIA)	
More effective management of customers.	Achieved
Increase in revenue through offering personalised services.	Not implemented
Increase in revenue through sales agreements	77% achieved
Increase in the number of partners.	Achieved
Increase in client's satisfaction and trust.	Almost achieved
Increase in partners' satisfaction and trust.	Almost achieved
DEMONSTRATOR #5 – PERSONAL DATA FOR MUNICIPAL SERVICES AND THE TOURISM INDUSTRY (PRATO)	
To increase the number of data owners involved	30
To increase the number of available data sources	2
To increase the number of shared datasets	70
To increase the number of activated smart contract	70
To increase the number of data analysis procedures	2
To increase the number of questionnaires/surveys	1
Improvement in the planning capabilities as perceived by the Office	4
Savings in the installation of traffic sensors and data acquisition procedures	The savings could be up

	to 100% in case a consistent number of users is involved
To increase the number of data owners involved	30
To increase the number of available data source	1
To increase the number of shared datasets	30
To increase the number of activated smart contract	0
To increase the number of data analysis procedures	1
To increase the number of questionnaires/surveys	1
Improvement in the planning capabilities of the cultural institutions	4
Savings in data acquisition and analysis procedures	No specific reduction can currently be established as the DataVaults tools work in addition to other tools
To increase the number of involved data owners	30
To increase the number of shared datasets	30
To increase the level of Prato administration's savings of resources in terms of costs and personnel	~ €0
To increase the level of Data owners' satisfaction in using the DataVaults tools	3.1
To increase the level of Data seekers' satisfaction in using the DataVaults tools	4.1
Common to all demonstrators	
Have we included personal datasets of 5 different sites/demonstrators?	Yes
Have we supported 20 types of personal data categories?	Yes
Is there compatibility with at least 20 types of data sources (sensors, IoT, APIs, wearables, records, etc.)?	No. 14 types
Are 12 known analytics algorithms supported?	Yes
Have we reused 10 existing vocabulary standards?	Yes
Did the five demonstrators successfully run for the required length of time?	Yes
Questions which can be asked of the DataVaults project for evaluation relating to strategic issues.	**Response**
Chapter 7 and deliverables such as D7.4 and D8.4 have covered many of these topics.	
Have we provided Policy Makers with faster and more effective decision-making procedures based on personal data?	Yes
Have we provided a solid reference implementation on which to base future legislation and regulations for personal data?	Yes
Have we made progress in advancing research and applying innovative technologies that utilise the best of breed in personal data management?	Yes
Has there been a positive effect of DataVaults on growing the Data Market?	Yes
Has there been a positive effect of DataVaults on growing the Data Economy?	Yes

Has there been a positive effect of DataVaults on Improving the Data Industry?	Yes
Has there been a positive effect on creating a growing eco-system?	Yes
Have we created a pan-EU platform?	Yes
Have we created a new citizen-centric data chain?	Yes
How have we contributed to EU Policy and Strategy?	Yes
In summary: Have we increased the size of the personal data lake? Has this contributed to the development and growth of an eco-system surrounding it?	Yes
Strategic Impacts: Have we contributed to the following?	
"Supporting the emergence of data markets and the data economy"	Yes
"… setting up and operating platforms for secure and controlled sharing of "closed data" (proprietary and/or personal data)"	Yes
"… address the necessary technical, organisational, legal and commercial aspects of data sharing/brokerage/trading, and build on existing computing platforms …"	Yes
"… preserve utility for data analysis and allow for the management of privacy / utility trade-offs, metadata privacy, including query privacy…"	Yes
Is there any evidence of an Increase in data economy activity?	Yes
Have we evidence of growing data eco-systems?	Yes
Can we show Financial & business impact in sectors working on personal data?	Yes
Have we influenced European policies on data protection and security?	Yes
Have we influenced other EU policies and strategies?	Yes
Have we made any contribution to standardisation?	Yes
Have we had other socially important impacts?	Yes
Have we enabled smaller players to participate in the data economy better?	Yes
Questions which can be asked of the DataVaults project to refine for evaluation relating to Non–Functional requirements	**Response**
Chapter 5 and the deliverables in WP1 and WP5 have covered many of these topics.	
Functional Suitability	
Is DataVaults able to collect data from Individuals in order to gather their data in one place? NFR1	Yes
Does DataVaults allow an Individual to select and manage how his/her data are to be shared to the DataVaults Cloud Platform? NFR2	Yes
Is DataVaults able to share the data collected from Individuals and make it available to Data Seekers following specific data sharing contracts? NFR3	Yes
Performance efficiency	
Does DataVaults guarantee the timely and robust collection of data from the side of the Individuals? NFR4	Yes
Is DataVaults able to handle and store datasets from various sources? NFR5	Yes
Does DataVaults guarantee the efficient and effective resource allocation for the sharing and encryption/decryption process execution NFR6	Yes
Is DataVaults able to perform analytics in a timely and efficient manner? NFR7	Yes
Does DataVaults guarantee the full optimization of the response time to ensure a functional and flexible navigation through the DataVaults solution? NFR8	Yes
Does DataVaults cater so that both the Public and the Private ledgers are able to process transactions fast and within certain time limits? NFR9	Yes

Does DataVaults provide prompt transaction responses from the Brokerage Engines? NFR10	Yes
Compatibility	
Is DataVaults able to interact and exchange information with other systems in a secure way (for example secure REST API)? NFR11	Yes
Does the DataVaults Cloud Platform provide communication capabilities to allow other applications to interact with DataVaults platform? NFR12	Yes
Does DataVaults allow the Personal DataVaults App to run on devices that do not support DAA? NFR13	Yes
Usability	
Does DataVaults feature a user-friendly interface, and be offering a set of user guides? NFR14	Yes
Does DataVaults provide a user interface that supports straightforward task accomplishment? NFR15	Yes
Does DataVaults provide the suitable error protection methods for all input fields? NFR16	Yes
Does DataVaults have a multi-language user interface? NFR17	Yes
Does DataVaults offer logs about evolution and faults history and periodically send debug reports? NFR18	Yes
Reliability	
Does DataVaults ensure high availability of the overall system? NFR20	Yes
Is DataVaults able to handle simultaneous requests on a timely and efficient manner? NFR21	Yes
Does DataVaults provide the mechanisms to recover the system state to normal operation after a failure? NFR22	Yes
Does DataVaults keep information about transactions parties private NFR23	Yes
Security	
Is DataVaults able to handle software errors without affecting the platform overall functionality? NFR24	Yes
Is DataVaults able to securely store uploaded Individuals' data? NFR25	Yes
Is DataVaults able to retain the privacy of Individuals based on the privacy level they have chosen? NFR26	Yes
Does DataVaults take into account privacy and security rules according to national legislation? NFR27	Yes
Does DataVaults ensure different authorisation access to different datasets? NFR28	Yes
Does DataVaults support data seeker's account validation? NFR29	Yes
Is DataVaults able to attest the identity of the user/subject performing any operation? NFR30	Yes
Does DataVaults provide the proper mechanisms for system upgrade/maintenance with minimum downtime? NFR31	Yes
Is DataVaults composed by components that are operating independently? NFR32	Yes
Is DataVaults able to raise alarms about hardware/software failures of the solution? NFR33	Yes
Does DataVaults provide strong transaction validation mechanisms? NFR34	Yes

Does DataVaults keep information about transactions encrypted? NFR35	Yes
Does DataVaults keep history of all important actions (such as transactions)? NFR36	Yes
Portability	
Is DataVaults able to be deployed in a timely and efficient manner? NFR37	Yes
Is DataVaults based on easily replaceable independent components interconnected through APIs? NFR38	Yes
Is DataVaults able to be deployed on various Linux based distributions? NFR39	Yes

APPENDIX 2, DATAVAULTS THEORY OF CHANGE

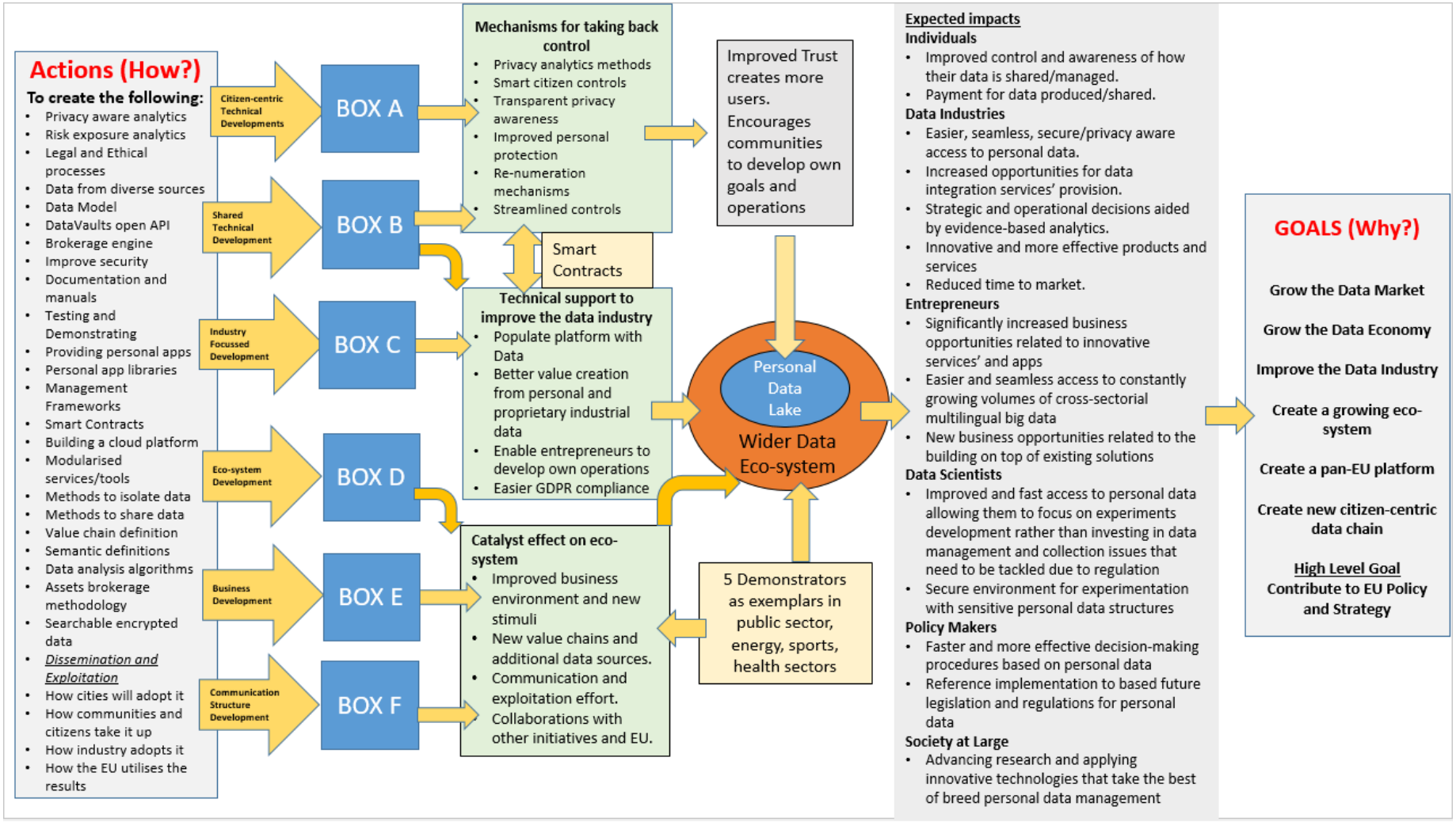

Figure 80: The DataVaults Theory of Change

BOX A: Citizen-centric Technical Developments
- Personal Data Management and Analytics App
- Open source library of the Personal Data App
- Privacy aware analytics
- Ethics monitoring framework
- Providing personal apps
- Personal app libraries
- Risk exposure analytics
- Methods to isolate data
- Methods to share data
- Legal and Ethical processes
- Risk Exposure Dashboard

BOX B: Shared Technical Developments
- Building a cloud platform
- Analysis on how personal data is/should be managed
- Improved security
- Smart Contracts
- Distributed ledger platform to facilitate transactions

Assumptions

The essential assumption is that DataVaults will increase the size of the personal data lake and as a consequence, this will lead to the growth of the eco-system around it to enable exploitation.
- 50% increase of data subjects on personal data platform
- 20% increase in data on the data platform
- 20% annual increase in the number of data provider organisations on the platform

There is cross-domain collaboration.
Business Innovation will create a new citizen-centric model and technical convergence.

BOX C: Industry Focussed Technical Development
- Data from diverse sources
- Data Model
- DataVaults open API
- Brokerage engine
- Management Frameworks
- Modularised services/tools
- Semantic definitions
- Data analysis algorithms
- Assets' brokerage methodology
- Searchable encrypted data
- Cloud-based Data Management/Analytics platform
- SaaS Platform,
- Brokerage platform expandable knowledge representation Secure cloud-based storage facility
- Cloud-based data analytics engine,
- Access control engine
- Risk management service
- Visualization library
- Personal data catalogue
- Business brokerage engine
- Platform architecture
- Designs of all software bundles
- technology support- security/ trust by design
- support for modern analytics algorithms on plain, multiplexed and encrypted data
- Data Access Framework
- Assets Brokerage Engine
- Secure Data Management Environment (cloud and App)

BOX D: Eco-system Development
- Documentation and manuals
- GDPR easier to implement
- Testing and Demonstrating
- Inclusion of personal datasets of 5 different sites/demonstrators
- Compatibility with at least 20 types of data sources (sensors, IoT, APIs, wearables, records, etc)
- 12 known analytics algorithms supported
- PESTLE analysis of the DataVaults ecosystem,
- SWOT analysis of the DataVaults platform,
- Needs elicitation survey

BOX E: Business Development
- Value chain definition
- Business Validation and Impact Assessment Report
- Different Business Model templates for personal data value sharing
- Consortium wide exploitation plan
- Financial strategy plan
- Consortium wide sustainability plan
- Individual partners business models and exploitation plans

BOX F: Communication Structure Development
- Dissemination and Exploitation
- Understanding how cities will adopt it
- Understanding how communities and citizens take it up
- Understanding how industry adopts it
- Understanding how the EU utilises the results
- Collaboration with at least 5 other projects and businesses handling personal data.

Figure 81: Key to the "Boxes" in previous figure