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

D1.1
DataVaults Data Value Chain Definition

Editor(s)	Tomas Pariente, Javier Villazán, Ricardo Ruiz
Lead Beneficiary	ATOS
Status	Draft
Version	1.0
Due Date	31/03/2020
Delivery Date	30/03/2020
Dissemination Level	PU

DataVaults is a project co-funded by the European Commission under the Horizon 2020 Programme (H2020-ICT-2019-2) under Grant Agreement No. 871755 and is contributing to the BDV-PPP of the European Commission.

Project	DataVaults – 871755
Work Package	WP1 - DataVaults Data Value Chain Elaboration and Concept Fusion
Deliverable	D1.1 - DataVaults Data Value Chain Definition
Editor(s)	ATOS - Tomas Pariente, Javier Villazán, Ricardo Ruiz
Contributor(s)	ASSENTIAN – Shaun Topham MIWenergia – Pablo Barrachina, Ramón Ruiz Olympiakos – Christina Tsiligkiri Prato – Paolo Boscolo, Elena Palmisano Piraeus – Michail Bourmpos SUITE5 – Sotiris Koussouris, Nefeli Bountouni
Reviewer(s)	UBITECH – Giannis Ledakis TECNALIA – María José López, Valentín Sánchez

Abstract	This document is the first deliverable of WP1 and presents the state-of-the -art on personal data platforms and their ecosystem. It also introduces the methodology to be followed till the completion of the work in WP1. In the document, the description of the demonstrator scenarios is presented and it shows the definition of the data value chain as well as the initial list of user stories from the demonstrators and a list of data sources per demonstrator that consists the base material upon which the project will work to deliver a platform of value for these demonstrators.

Executive Summary

This document provides an overview of the state-of-play of personal data platforms and goes on to record the initial usage scenarios of the demonstrators of DataVaults, with the aim to sketch the high level picture of the data value chain within DataVaults and with the definition of an initial set of user stories, alongside with required data sources that would allow the demonstrators' scenarios to be realised.

At start, the document at hand goes into the details of existing efforts and tools dealing with personal data management and their methodologies. It also provides an overview of different approaches taken by different stakeholders involved towards a personal data economy, looking at different angles and potential revenue streams behind data platforms.

Following, the document looks to the existing demonstrators within the project to identify possible scenarios of usage, deriving from them a set of user stories according to a methodological approach proposed.

Last but not least, the document provides and initial catalogue of data sources used by the demonstrators, as well as a wish list of potential data sources of interest by the demonstrators to be used to complete the proposed scenarios. As data availability is a factor that may change from time to time, based on the update of devices APIs, the introduction of new APIs and data sources, the terms of use of the various services, as well as legislation, list will be revised during the WP1 tasks, in order to include all data sources relevant to the final demonstrator scenarios, as well as other data sources that will provide added value to the overall DataVaults platform.

This deliverable from WP1 is the first one setting the scene for the project development of the personal data platform. An updated version of the document will be provided in M18 (June 2021). The document, along with the deliverables produced in WP2, will be used also as one of the inputs to the implementation WPs (WP3 to WP6).

Table of Contents

1 Introduction 6
1.1 Motivation 6
1.2 Document structure 7
2 State-of-the-art of Personal data platforms 9
2.1 Personal data platforms stakeholders 9
2.2 Personal data management platforms 10
2.3 Personal data valuation methodologies and economy 15
3 Demonstrators scenarios 18
3.1 Sports and activity personal data 18
3.2 Strengthening entrepreneurship and mobility 19
3.3 Healthcare data retention and sharing 22
3.4 Smarthome personal energy data 24
3.5 Personal data for municipal services and the tourism industry 26
4 Personal Data value chains 29
4.1 Demonstrators personal data value chains 30
4.2 Initial Demonstrator Requirements 33
4.2.1 User stories 34
4.2.2 User Stories provided by the demonstrators for identification of technical requirements 35
5 Data sources 44
5.1.1 Sports and activity personal data 44
5.1.2 Strengthening entrepreneurship and mobility 45
5.1.3 Healthcare data retention and sharing 46
5.1.4 Smarthome personal energy data 46
5.1.5 Personal data for municipal services and the tourism industry 47
6 Conclusion and Future work 49
7 References 50

List of Figures

Figure 1 – BDVA Reference Model and DataVaults category marked 6
Figure 2 – NIST Big Data Reference Architecture [3] 7
Figure 3 - Data stakeholders. Source: [5] 10
Figure 4 - Possible use fees 16
Figure 5 - Aerial view of Piraeus 20
Figure 6 - Screenshot of PireApp 21
Figure 7 - Andaman7 application 22
Figure 8 - DataVaults Smarthome 24
Figure 9 - Satisfaction analysis and tourist services in Prato 27
Figure 10 - General data value chain 29
Figure 11 - Methodology to define the data value chain 30
Figure 12 - User sharing Data Value Chain 31
Figure 13 - Data seeker Data Value Chain 32
Figure 14 - Core DataVaults data value chain 33

List of Tables

Table 1 - Existing personal data platforms 14
Table 2 - User stories provided by the Demonstrators 43
Table 3 - Olympiacos Available data 45
Table 4 - Olympiacos Needed data 45
Table 5 - Piraeus Available data 45
Table 6 - Piraeus Needed data 46
Table 7 - Andaman7 Available data 46
Table 8 - Andaman7 Needed data 46
Table 9 - MIWEnergia Available data 47
Table 10 - MIWEnergia Needed data 47
Table 11 - Prato Available data 47
Table 12 - Prato Needed data 48

Terms and Abbreviations

PdP	Personal data platform
PV	Photovoltaic
ROI	Return of investment
BDVA	Big Data Value Association
SRIA	Strategic Research and Innovation Agenda
NIST	National Institure for Standards and Technology
NBDRA	NIST Big Data Reference Architecture

1 INTRODUCTION

1.1 MOTIVATION

The main goal of this document is to provide a state-of-play view of personal data platforms, including existing trends on personal data management, tools and methodologies. The document provides also an overview of how DataVaults may tackle the different aspects of a personal data value chain.

The data value chain is rooted in previous work in industry where a value chain was defined as a collection of steps to be followed to deliver a product to the market [1]. This concept of a value chain was adopted to handle the different steps to deliver a data-intensive product or application, giving birth to the so-called data value chains. A data value chain consists in several steps, typically related to the phases of data acquisition or collection, storage, analysis and usage.

There are several attempts to standardize or formalize data value chains, taking into account not only pure aspects of data management, but many other related issues, such as security, privacy, etc. Of special interest for DataVaults is the BDVA Reference Model provided in the Strategic Research and Innovation Agenda (SRIA) [2] by the Big Data Value Association (BDVA). Figure 1 shows the BDVA Reference Model.

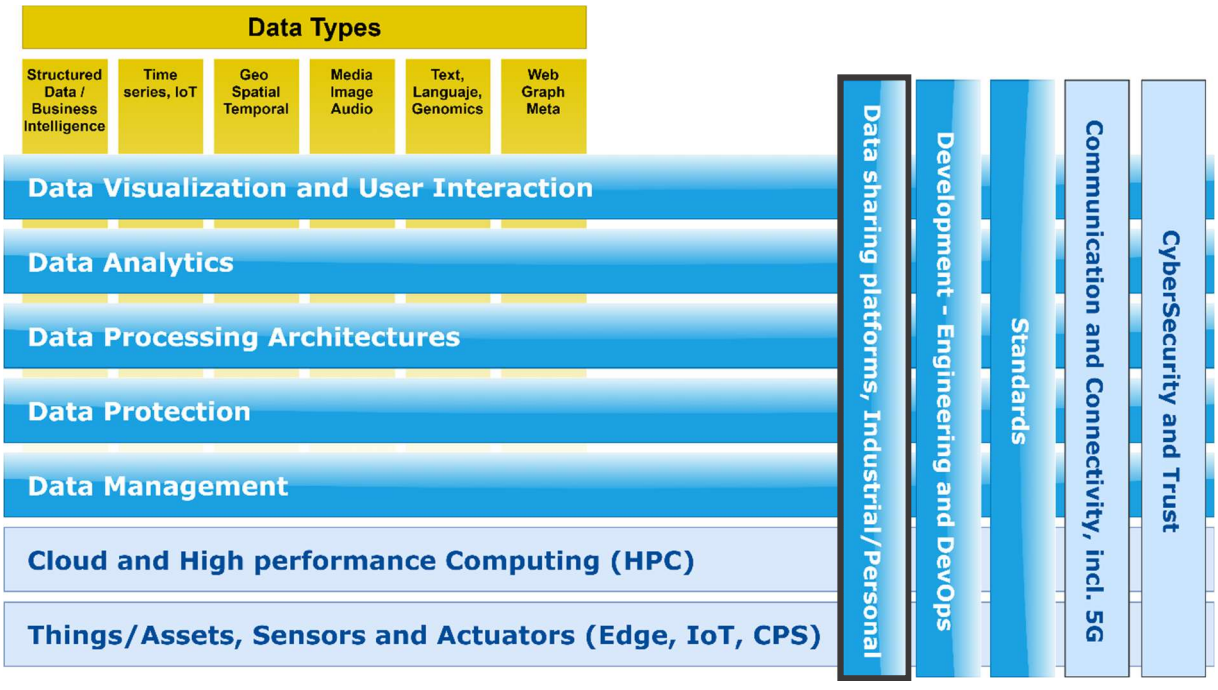

Figure 1 – BDVA Reference Model and DataVaults category marked

The BDVA Reference Model, is important for DataVaults, as our project is part of the portfolio of projects funded under the Big Data Value PPP umbrella. Therefore, it is important to align our work to this reference model where possible. In this model, the data value chain goes in different axes. The bottom-up axis represents different concerns related to data, such as the collection (for instance from IoT devices), to the processing infrastructure (HPC, Cloud or Edge), but also aspects related to data management, data protection, data processing,

analytics and usage. The vertical yellow lines represent different data types that require usually different type of tools and methods from the data value chain. DataVaults belongs to the data platforms represented by the "Data sharing platforms, personal/industrial" located in blue in the vertical axis.

Another important reference related to data value chains to take into account is the NIST Big Data Reference Architecture [3] developed by NIST, the US institute for standards in technology. Figure 2 represents an overview of the NBDRA.

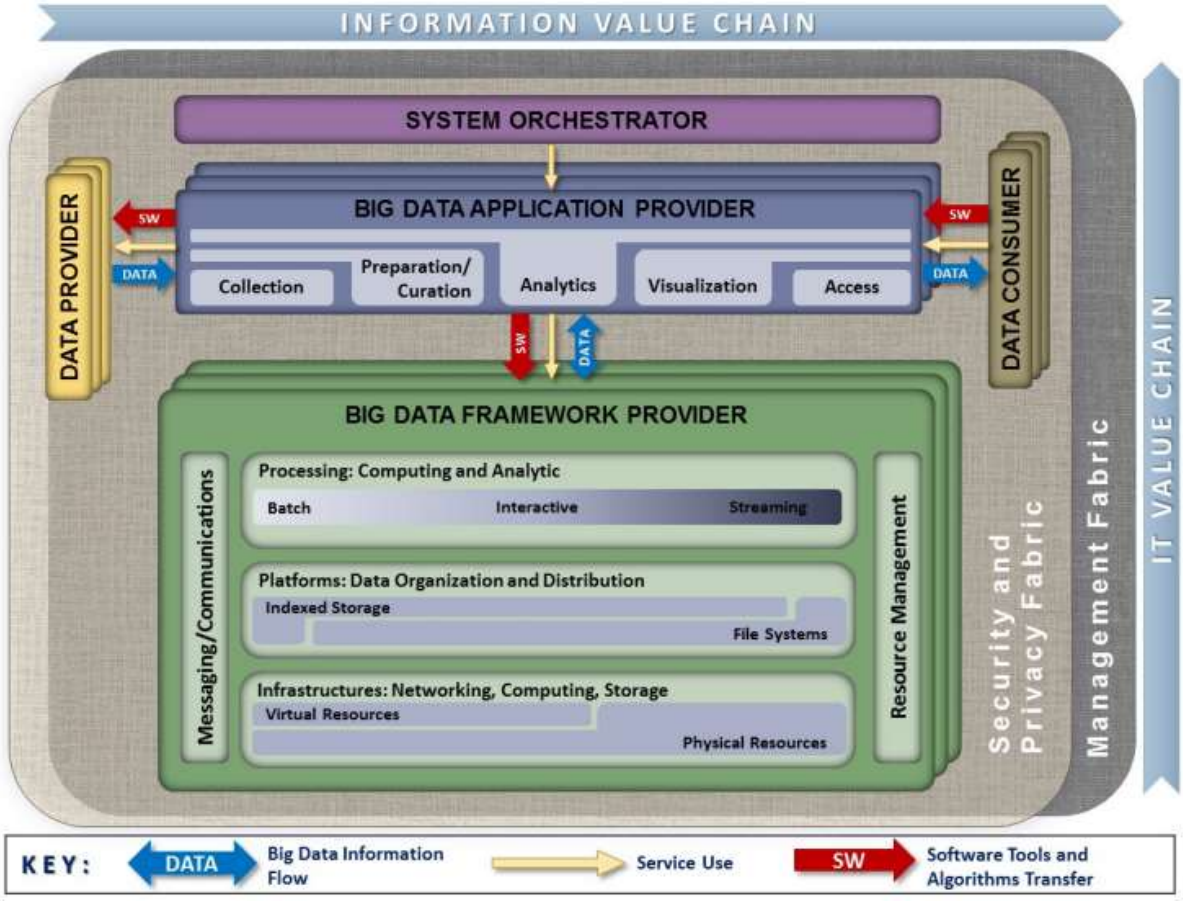

Figure 2 – NIST Big Data Reference Architecture [3]

The NBDRA represents an attempt to standardize big data architectures. For DataVaults it is interesting the so-called "Information Value Chain" that goes from left to right and represents the big data value chain, with steps such as data collection, preparation/curation, analytics, visualization and access. This is in line with the concept of data value chain that we are following in the project. Moreover, the "Security and Privacy" layer that surrounds the main architecture is related to the concept of personal data platforms.

1.2 DOCUMENT STRUCTURE

This document is structured as follows:

Section 1 introduces the document.

Section 2 analyses and describes the current State-of-the-art in data platforms and the data sharing domain. Starting with the stakeholders involved in the data value chain in Section 2.1,

moving then to current data platforms in Section 2.2 and finalizing with the methodologies an economy behind the data ecosystem in Section 2.3.

Section 3 describes the different scenarios planned by the demonstrators involved in the project. It states the current problem that the demonstrators are facing, then goes into a description on possible scenarios that could improve that problem. It provides also the description of the demonstrator after the project is implemented.

Section 4 provides the methodology and the data value chain for the project. An iterative methodology is presented in order to identify and describe the project's way of working regarding the data value chain. The section describes the core data value chain concept for the project. Moreover, in Section 4.2 the document introduces the methodology to derive the user stories from the scenarios and the list of identified user stories.

The Section 5 of the document lists and describes the identified data sources used by the demonstrators and also the data sources identified as possible data sources to be used for some of the scenarios.

The document concludes with the most important remarks extracted from this document and the description of the future work planned for the next steps of the project in Section 6.

2 STATE-OF-THE-ART OF PERSONAL DATA PLATFORMS

This section aims to present the state-of-the-art for personal data storage.

The content of this section is split in three parts. First, we introduce the different stakeholders identified in the literature and the data platforms ecosystem.

Then, we move on to the state-of-the-art description of the ecosystem for personal data platforms. A table with a list of already existing personal data platforms as well as its description and evaluation are described in this subsection.

Finally, we will go through the literature to describe the current state of the personal data economy and how the data is being valued and shared in this type of platforms. Moreover, we will describe the methodology used in the current literature to define the data and the platforms.

2.1 PERSONAL DATA PLATFORMS STAKEHOLDERS

Taking a look at the stakeholders involved in the personal data platforms consists in understanding the interests, and therefore the benefits and provenance of that data.

As mentioned in [4], people (individuals) are the beginning and the end of the personal data value chain. We are the ones that generate the data, and we are the final ones that consume them (directly as data, or indirectly as services coming out infrastructures/providers who process these data and use them in their service offerings).

Individuals are the first ones interested in their own data, but as part of a society, that data is also interesting as a small part of a bigger combination used for analysis. Living in a capitalist society where the "average" mathematical term is so important, the comparison seems to be also important. Such comparison is only possible with the aggregation and combination of data from as many users as possible.

Therefore, we have the "who" and the possible "why". With these two terms it will be easier to identify other stakeholders in the personal data and understand why this flow is fed back.

The table represented in Figure 3, obtained from [5], gives a good view of the stakeholders involved in the personal data value chain.

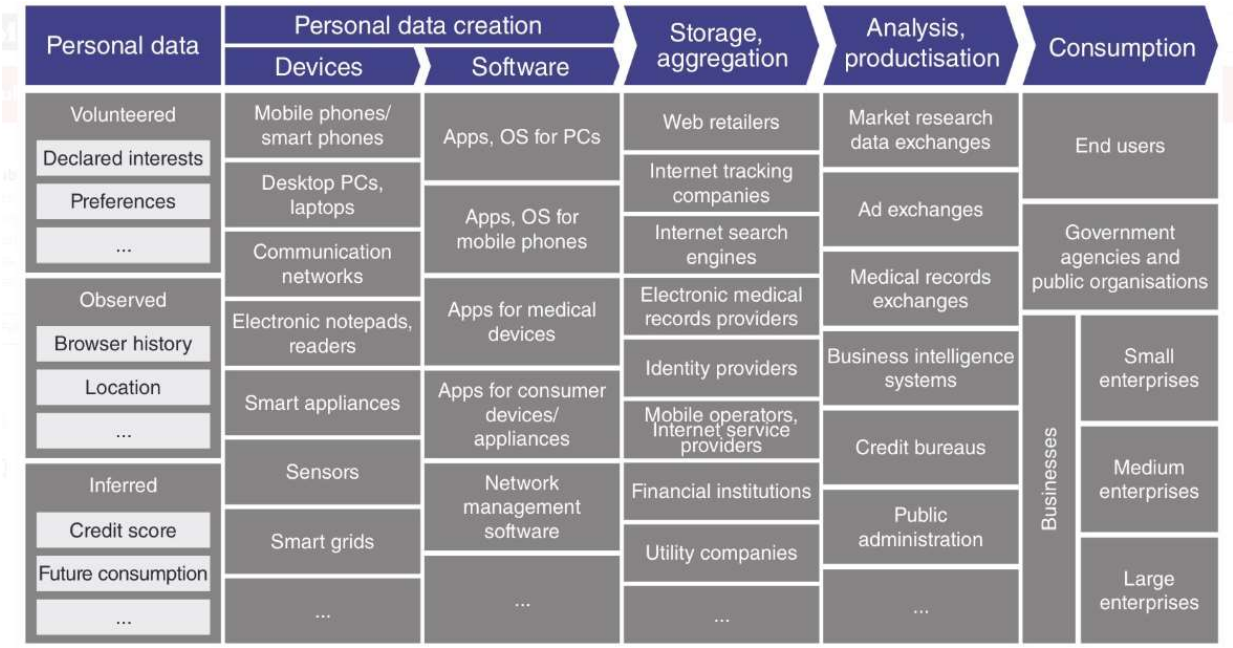

Figure 3 - Data stakeholders. Source: [5]

From the column of "**Personal data**", we can retrieve that the data could be provides as:

- Volunteered
- Overserved
- Inferred

As "volunteered" data we can assume that we generate data which we allow to share it if we want, while the "observed" and the "inferred" data might generate legal issues, as we really don't know who the owner of it are. But this gives a good approach to the personal data and the legal issues that such data might generate.

Many international or national companies can be identified from the columns of "**Personal data creation**", "**Storage aggregation**" and "**Analysis, productization**" that play those roles, which are specialized in certain domains to make the personal data profitable for themselves.

"**Consumption**" is committed to our first idea that the main consumer of our personal data is our self, but there are other consumers of these data related to the public sector, such as Treasury, Police, Justice departments, etc. as well as a lot of enterprise-related with the private sector. Some of the stakeholders the table does not describe are the legal bureau and consultants. These could have their personal column related to the legal issues involved in the use of personal data. These stakeholders are the ones who advise if the personal data of a third party can be used, or when your personal data are used in a legal or illegally. In a certain way, they could be called "The data guards" that act on their own or ordered by someone that feels that their personal data is used in a bad way.

2.2 PERSONAL DATA MANAGEMENT PLATFORMS

One of the targets of this section is to take a look at the existing apps or platforms that are involved in the management of the personal data of the people. Taking into account that all these apps or platforms are centred in giving the management to the individuals, which are

the ones who decide whether or not to share the personal data, with whom and at which moment.

Some of these apps or platforms are focused on sharing a huge different kind of personal data, such as the social, financial, etc but other are more focused on a specific kind of data, such as the medical personal data.

The table below gives a description of the most relevant apps or platforms found.

APP NAME	DESCRIPTION	SUMMARY	PERSONAL DATA USABILITY APP PROVENANCE	MONETARY MODEL URL
DIGI	Digi.me is an app that allows you to manage all your personal data with a new model for privacy working toward a decentralised future with proven reliability, privacy and security today	• Data is encrypted inside the digi.me Private Sharing app as it imports from the individuals accounts • The encrypted data is stored in a personal cloud (Dropbox, Google Drive, Microsoft OneDrive) • Data is encrypted with a set of keys and ciphers equal to those of banks and the military • Data can only be viewed and privately shared when the individual is logged in and give explicit permission • All data operations happen inside the app or within a temporary virtual personal cloud (for syncing) that terminates when done • DIFI.me never see or store the client password • Private Sharing with apps is an exciting new way to power digital services while protecting your data and privacy • The Private Sharing consent screen clearly explains what data will be used, for what purpose, and whether a copy will be stored • Log of the private shares inside digi.me • Digi.me makes money from small transaction fees every time you use Private Sharing to power an app	Social app, Bank, Medical, Sports app, Personal data UK, USA, AUSTRALIA BOSNIA & HERZEGOVINA	FREE (for individual) PAYMENT (for business) https://digi.me/
BALI	Bali enables users to aggregate their information securely and privately and use it for their own benefit	• Bali enables users to aggregate their information securely and privately and use it for their own benefit.	A wide range of personal data USA	FREE https://www.microsoft.com/en-us/research/project/bali/
HAT	The HAT Microserver is a new, fully scalable and advanced technology that confers intellectual property rights of personal data to individuals through their ownership of a dedicated database, wrapped with containerised microservices	• Technology that can enable individuals to own our own data and yet be able to interact with all the websites and apps that we love • Reuse and re-share this data whenever we wish, • The HAT microserver is fully portable across devices, but is commonly hosted in the cloud • The HAT is fully open sourced but services in the HAT ecosystem are built by commercial as well as non-profit organisations.	Words, photos, music, locations, financial transactions, etc. UK (Cambridge)	FREE https://www.hubofallthings.com/
MYDEX	Mydex is a platform that gives organisations and individuals the tools they need to interact over the	• Individuals can store, manage and share a wide range of personal data, both verified and self-asserted with our data management API. This data can be combined with their MydexID to	A wide range of personal data	FREE (for individual) PAYMENT (for business)

APP NAME	DESCRIPTION	SUMMARY	PERSONAL DATA USABILITY APP PROVENANCE	MONETARY MODEL URL
	web. It's safe, secure and cost-effective. Individuals can store, manage and share a wide range of personal data.	prove who they are, their entitlement to services — to get things done online • Mydex uses Amazon Web Services as its hosting provider. All your data is stored in encrypted form on UK-based servers managed by a specialist provider whose reputation has been built on ensuring the privacy, safety, and security of commercial and public sector customers, government, health and education around the globe. • The platform is constantly monitored and rigorously controlled using 'zero touch' deployment tools, static code analysis, code hardening and a hybrid testing model. • Encrypted data, encrypted communication, one-time password which is delivered via what is called a asymmetric key pairs, these are unique to your connection with that organisation. Asymmetric keys enable data to be verified in terms of who the sender is data encrypted with one key can only be decrypted by the other key. (https://dev.mydex.org/fyi/security-model.html)	UK	https://digi.me/
OWNYOURINFO	OwnYourInfo is an app that puts you in control of your data and allows you to keep your important information and files safe, organized and up to date. It also gives the possibility to share your information securely with anyone from anywhere.	• Caregivers and parents use OwnYourInfo as a means to manage and share their dependents medical information. MEDIAL INFOR • Individuals concerned about their growing information get organized using OwnYourInfo to store health, financial, passwords and other critical information • Travelers use OwnYourInfo to keep vital information with them on-the-go, e.g. a copy of passport, drivers' license and birth certificate. • Free or Premium. Free Single Profile "Information Templates" "25 Encrypted Files" "5 Shares / Month". Premium profile "30 GB File Storage" "Unlimited Shares" "Information Reports" 5$/month • OwnYourInfo's secure sharing solution enables a private environment for the most secure information exchanges.	Health, financial, passwords and other critical information USA	FREE & PAYMENT http://www.ownyourinfo.com/
MECCO	Meeco gives people and organisations the tools to access, control and create mutual value from personal data in a privately, securely and with an explicit consent.	• ENTERPRISE -- Enable customers to control their personal data, manage data risk and compliance, Build trust and boost loyalty • FOR EVERYONE -- Securely store your personal data with access across all devices, connect with the people and organisations you trust, edit once, automatically update trusted contacts. • Meeco API, Meeco LABS.	Personal data (depends on the services integrated by third parties) AUSTRALIA	FREE for users Not described for enterprises https://www.meeco.me/
PEOPLE	people.io is an app that allows you to connect to other apps to provide you a better and more interesting ad content while ensuring you retain control of your data.	• Allows a securely connect accounts from across the personal digital life and they will use their proprietary machine learning algorithms to create a single view of the person. • The app always has an eye on the privacy and security of their community, applying industry-leading techniques to safeguard their users.	A wide range of personal data (Depends on the country) ENGLAND AND WALES	Not Described http://www.people.io/

APP NAME	DESCRIPTION	SUMMARY	PERSONAL DATA USABILITY APP PROVENANCE	MONETARY MODEL URL
		• people.io app allows to connect to other apps and they'll help them enrich the people's experience by tailoring content and recommendations.		
CITIZEN ME	CitizenMe is a platform that gives you access to unique data points around the world.	• USE CASES, interesting because is near to real life • Templates for surveys • Have a way to ensure that the people are real, has advanced verification methods • All users' data is stored on their device and backed up to a personal cloud. When they choose to exchange it with an organisation, a copy of the requested data leaves the user's' device in an anonymised data file.	A wide range of personal data UK	PAYMENT https://www.citizenme.com/public/wp/human-data/personal-data/
SOLID	Solid (derived from "social linked data") is a proposed set of conventions and tools for building decentralized social applications based on Linked Data principles. Solid is modular and extensible and it relies as much as possible on existing W3C standards and protocols	• Solid (derived from "social linked data") is a proposed set of conventions and tools for building decentralized social applications based on Linked Data principles. Solid is modular and extensible and it relies as much as possible on existing W3C standards and protocols. • Angular	SOCIAL USA	FREE https://solid.mit.edu/
PEERCRAFT	Peercraft makes it easier and safer for consumers to make good deals with both companies and private sellers. We start with a basic service, which will gradually be developed in dialogue with the users.	• Peercraft makes it easier and safer for consumers to make good deals with both companies and private sellers. We start with a basic service, which will gradually be developed in dialogue with the users. • Access to websites with a simple but secure login • Control which information you provide to services • Enter into agreements with companies and individuals via Peercraft (Coming) • Locate businesses offering specific goods or services (Coming) • Check and report the vendor's qualifications and reputation (Coming)	A wide range of personal data DENMARK	FREE https://www.peercraft.com/
OPENPDS/SAFEANSWERS	openPDS/SafeAnswers allows users to collect, store, and give fine-grained access to their data all while protecting their privacy.	• openPDS/SafeAnswers allows users to collect, store, and give fine-grained access to their data all while protecting their privacy. • Pandora	A wide range of personal data USA	FREE https://openpds.media.mit.edu/
COZY	Cozy is a Smart personal cloud to gather all your data.	• Cozy Cloud offers everyone a smart digital home, Cozy that combines comfort and security for more services. • FREE and PAYMENT depending on the amout of data in GB	No Restriction FRANCE	FREE AND PAYMENT https://cozy.io/en/
MIDATA	MIDATA may actively contribute to medical		Medical	FREE (Cooperative)

APP NAME	DESCRIPTION	SUMMARY	PERSONAL DATA USABILITY APP PROVENANCE	MONETARY MODEL URL
	research and clinical studies by granting selective access to their personal data	• It has a clear idea about the legal issues of the data, ethics and governance rules, https://www.midata.coop/en/partners/.	SWITZERLAND	https://www.midata.coop/en/home/
MYHEALTHMYDATA	MyHealthMyData (MHMD) is a Horizon 2020 Research and Innovation Action which aims at fundamentally changing the way sensitive data are shared. MHMD is poised to be the first open biomedical information network centred on the connection between organisations and individuals, encouraging hospitals to start making anonymised data available for open research, while prompting citizens to become the ultimate owners and controllers of their health data	• BLOCKCHAIN • SECURE COMPUTATION • DISTRIBUTED LEARNING • SYNTHETIC DATA	Medical EUROPEAN PROJECT	FREE http://www.myhealthmydata.eu/

Table 1 - Existing personal data platforms

The table above, Table 1, gives a view of the most relevant platforms found based on personal data. Hosted all over the world, the perception is that personal data is not only important in a specific part of the world and these platforms are not only interested in the individual's personal data of the place where they are hosted.

These platforms give different solutions to store data using the Cloud, store by themselves solutions or aiming to the individuals to store their own data in their different devices. Even though we are at the very begging of the project, the first idea for DataVaults will be based on a Cloud solution to store data.

As we have mentioned in other sections, personal data can be used in different scenarios that can be profitable for the individuals but also for the enterprises. This could be the reason why these platforms became so important in the last times and is because of this why some of these platforms are payment applications in some cases.

The aim of these platforms is to give control of personal data to the individual or to the company which stores it. DataVaults goes in the same direction but also considering the importance of the legal issues and privacy. This will be one of the challenges that the project has to achieve in the future.

2.3 PERSONAL DATA VALUATION METHODOLOGIES AND ECONOMY

In this section we will try to identify and analyse potential methodologies for determining the economy of personal data. Most of the current literature agrees that there is no commonly accepted methodology for estimating the real value of personal data.

In "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value" [6] the author identifies 6 possible methodologies to be used in this determination.

They categorize them in 2 big groups:

- **Based on market valuation:**

  They identify those that refer to values that can be gathered or derived from the markets such as: "market revenue income per data record, market price for data, cost of data breach and price in illegal markets".

- **Based on individuals' valuation:**

  They split them in the results from surveys and willingness of users to pay to protect their data.

As depicted from the document, none of these methodologies are perfect and all of them have flaws that prevent from gathering a common value of the personal data. On top of that, these data valuations are mostly focused on US market where the biggest data firms are based, and that value might not be shared cross-regionally.

Other research try to focus on quantifying the data value by analysing the cost of "data breaches" in the companies [7] [8] concluding that the number of data breaches is increasing and also the impact and the cost. But it is not straightforward to quantify the value of the data by only looking at the data breaches.

Moreover, the constantly changing environment and legislation regarding personal data makes it even more difficult to quantify.

In this uncertain environment, the most logical way of analysing the value of personal data is by asking the involved stakeholders about the value that they give to that data. One way of doing so is by setting a cost of the data transactions instead of the data itself.

In the previous documents they identify some possible fees that can be used to identify the cost of the data movement. They are summarized in the Figure 4.

They can be categorized in 3 big groups:

- **Connection fees:** Fee that is payed upon connection to the platform, every time a stakeholder interacts with the platform to store or extract data, they need to pay a fee. It could be something similar to the approach that some online newspapers are following, where you can pay a fee to see a single article instead of the whole membership.
- **Membership fee:** This is the approach followed by many online services where the stakeholder needs to pay a membership before actually being able to use the service.

In some cases, it is shown as freemium service where the service can be used for free with certain limitations.

- **Data transaction fee:** This is the fee that is payed per transaction. The perfect example is the cloud services where they pay-as-you-use way of fee is implanted. In this case, given that it is a data platform it is logical to think as a cost per data unit transferred from o to the platform.

In this project we explore the idea of a possible extra fee, the one that is payed by the data seekers directly to the users for their data, which is shown in Figure 4 as "Reward for using their data". In this case the platform acts as a trusted intermediary that validates the contract, but the payment is not done to the platform but to the user directly.

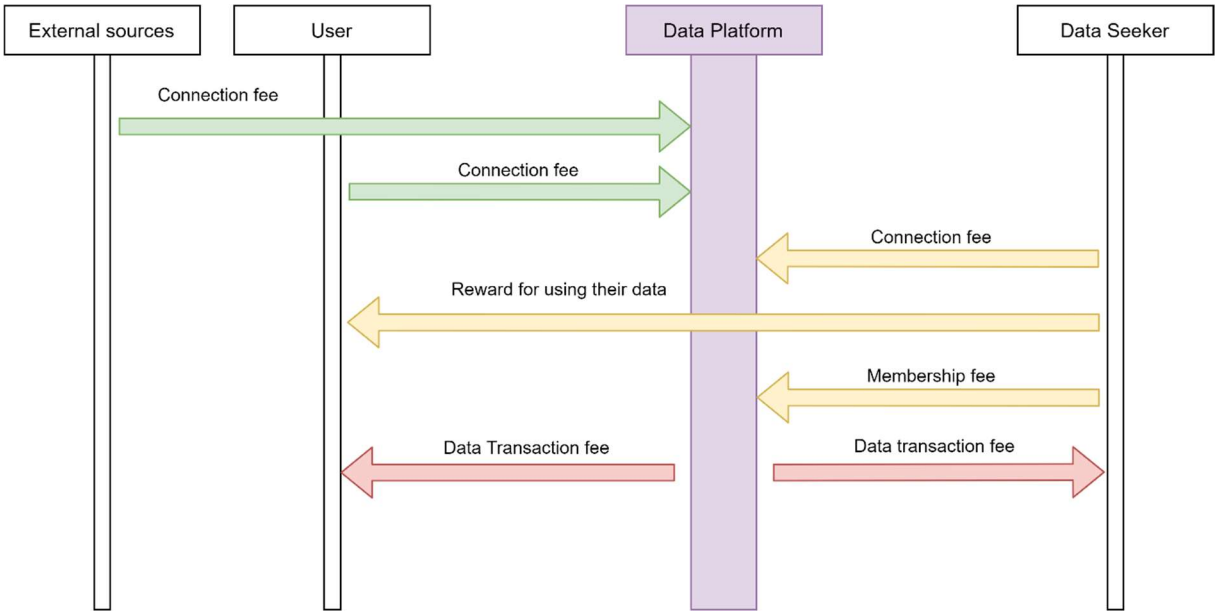

Figure 4 - Possible use fees

In "Paying for privacy and the personal data economy" [9], the author introduces the concept of Pay-for-Privacy as an iteration over the traditional data models where the data was used as payment in a "freemium" concept. In this document they identify the increasing willingness of users to pay for their data privacy and data security mostly related to the increase of data breaches and the impact that they have.

Most of the documents read, generate a different vision of the value of personal data, so give a unique vision of how the personal data can have a value or an economic impact is a mess and could not cover the whole impact of this particular project. Having this uncertain environment, the most logical way of analysing the value of personal data is by asking the involved stakeholders in the DataVaults project. Each of these stakeholders should probably have different ways to give value to the personal data that they have. But these data are not only profiting the stakeholders they also profiting to the ones that are sharing it, in this particular case we are thinking in the people that share the data, which also can benefit from this value chain. In some way, the chain seems not to be linear and we can identify it like a cycle that is constantly feeding back, being a benefit for all the players of the chain and in the end, being a benefit for society.

We have identified three different roles in this chain, the one refers to private enterprise, public institutes (university, etc) and the one refers to the people. Understanding that these roles can have different needs we can approach in which way they want to give value to the data.

3 DEMONSTRATORS SCENARIOS

In the following sections we introduce the high-level description of some scenarios, already determined by the DataVaults demonstrator partners, that fit best with the project's scope.

In this deliverable we showcase a number of diverse scenarios per demonstrator. In the scope of the project we expect these scenarios to be modified, merged or even discarded because of the value that they add to the project is not enough. But, in order to not being constrained to some early decision that could impact the rest of the project, we have listed and describe here all the identified options.

3.1 SPORTS AND ACTIVITY PERSONAL DATA

PROBLEM STATEMENT

OLYMPIACOS has a large base of 130,000 registered members and 70,000 active fans who register or renew their subscription on an annual basis. This base and the satisfaction of the members and fans are very important issues for the club as they are the most important source of revenues. Moreover, as the club has eighteen different sport departments at the competitive level and the Academies, needs to manage a large base of 2,000 professional and young athletes including important contact and medical details. A crucial issue for the club is the management of the data included in these two bases; on the one hand, the data included in the base of members and fans are compulsory so as to organize and coordinate procedures such as the organization and the participation of the members in the club General Assembly, the right to vote in the Administrative Elections and take part in the decision-making. Moreover, this database is also beneficial for the planning and the implementation of segmented and targeted marketing campaigns, the improvement of fan engagement and finding sponsors. This database concerns personal information, demographics (e.g. age, gender, nationality, address, postal code etc.), contact details (e.g. phone number, email address etc.), personal governmental data (e.g. Insurance number), personal social media accounts (e.g. Facebook, Twitter, etc.). For the club, it would be extremely beneficial to collect information such as user preferences, activity and mobility data in the future so as to focus on the real preferences of the members and fans and improve the services and the fan engagement in order to ensure their loyalty. The athletes' base requires better management of the results of the ergometric and medical examinations and statistical reports regarding their performances in training and the matches. This base also includes personal information, demographics, and sport exercise/activity one. Better management is very crucial for the club to adapt and plan the training sections and the team tactics, make players' transfers with specific physical skills, cover the athletes' expectations offering the appropriate medical and sport equipment. Regarding the data collection of the members' and fans' base, the club uses a system that offers the chance to register and renew their subscription on their own via the official club website (olympiacossfp.gr). In this case, an important issue is that a number of fans are hesitant to give all their data because they concern about data protection. The CRM tools export incomplete reports making the implementation of successful strategic and marketing plan difficult. As far as the data collection of the athletes is concerned, this is a manual procedure. The club has also started to use a CRM tool for the data management of

the athletes. A critical issue is that the version of the CRM that the club uses, doesn't allow the entry of more complicated data such as the results of the ergometric and medical examinations. So, there is a need for an appropriate platform that facilitates the entry of more complicated information.

CONSIDERED SCENARIOS

- **Scenario 1: Members/Fans – Key stakeholders/sponsors:** With the development of this scenario we encourage fans and athletes to capture, witness and manage their personal data through the platform so we can create a stronger relationship/interaction with the fans and encourage to care about the sharing, availability and keeping the data up to date.
  This could help Olympiacos to pursue the following step: Reorganization of the strategic plan, respond to the requests that may come from different organizations such as sponsors/NGOs/Federations/Local authorities who want to run a campaign/host an event for the club members/fans and athletes and finding new sponsors/partnerships
- **Scenario 2: Athletes:** Change in training sections, team tactics that implies covering the athlete's expectations offering the appropriate medical and sport equipment

DataVaults offering to demonstrators

OLYMPIACOS aims to adopt DataVaults, use it and promote it to the key stakeholders (members, fans, athletes, sponsors) to develop a stronger relationship with them and encourage them to care about the sharing, availability and update of their data (especially those are related to social media activity and the preferences) securing their privacy.

This would assist the club to re-use the current data and use new ones to engage and interact with the key stakeholders better and reorganize its strategic plan (new market segmentation, marketing campaigns for specific target groups, finding specific sponsors etc.) aiming to the improvement of the services and the increase of the revenues.

3.2 STRENGTHENING ENTREPRENEURSHIP AND MOBILITY

PROBLEM STATEMENT

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 center of the Athens metropolitan area, and one of the most densely populated cities in Europe (over 180.000 inhabitants, 15.065 citizens/km2) based on the 2011 population report. Apart from commercial trading, the traffic of cruise passengers is very remarkable (over 12 million travellers in 2018), causing equally **heavy traffic** in the road network of the Piraeus centre (40.350 daily moves).

Figure 5 - Aerial view of Piraeus

Aside from its importance as a tourist destination and a starting point for hopping onto cruises, Piraeus has one of the biggest commercial areas in Greece, having a large market that is based both on local citizens, as well as on tourists. However, the recent **financial crisis** has also hit this market, resulting in a decrease of new business ventures of around 81,6% compared to the growth witnessed 10 years ago. In order to counterbalance this deficit, and to restart the local economy, the city of Piraeus got engaged in Digital Strategy actions, while an Integrated Territorial Investment program of the amount of 80M€ has been issued, to make the city more attractive both to citizens and visitors, offering amongst others novel ICT services.

One of those is the smart phone application PireApp, aiming to boost citizen engagement. Though the use of this app, citizens and visitors are able to get in direct communication with the local authorities and report problems associated to urban the environment, while at the same time a database of profiles is generated, helping the city to better identify its citizens and their needs. However, **the information recorded through the app is still limited** and does not allow the municipality to perform complex analyses that could provide insights to help the city understand and support local entrepreneurship.

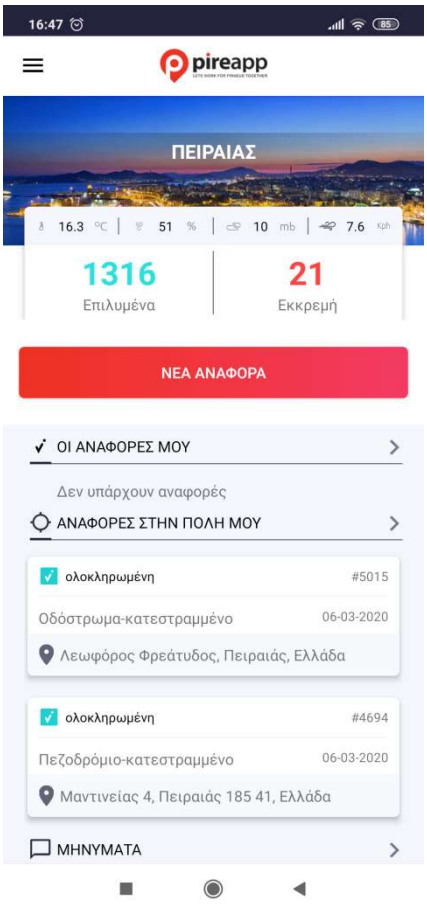

Figure 6 - Screenshot of PireApp

CONSIDERED SCENARIOS

- **Scenario A. Smart Mobility Services for Individuals**. This scenario will engage both OLYMPIACOS and PIRAEUS and will use the data shared by the interested citizens, as well as by the members taking part in the OLYMPIACOS demonstrator, to better schedule the mobility strategy and the relevant services within the city. The specific area of interest during the course of the project will be the surroundings of the OLYMPIACOS sport venues.
- **Scenario B. Empowering local entrepreneurship**. 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. Moreover, PIRAEUS will invite local entrepreneurship associations (i.e. the Piraeus Traders Association) and other interested stakeholders to either join the platform or act as 2nd tier data seekers, to test the aspects of the project that have to do with value generation and sharing with entities not directly using personal data but that access the derivatives of the latter. This scenario meets the on-going activities of PIRAEUS about the city's Open Trade Centre associated, inter alia, to the improvement of the local economy through restructuring of the market infrastructures and the deployment of smart applications.
- **Scenario C. Services for Personalised cultural and touristic experiences**. This scenario will build 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 to be analysed will 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. This scenario is both aligned and complementary to the "Digital Strategy 20" of PIRAEUS in terms of implementing an integrated Destination Management System, engaging citizens and visitors in the interactive definition of the cultural content of interest through the analysis of public (i.e. museums & touristic organisations) and private (i.e. travel agencies, cruise operators, booking organisations, etc.) data sources.

DataVaults offering to demonstrators

Through the DataVaults project the Municipality of Piraeus aims to provide better mobility services for individuals, when big events occur in the Olympiacos' stadiums. Also, through targeted joint actions from the Municipality and local stakeholders we expect to enhance the local market and promote local entrepreneurship. Finally, through the use of personal data analysed form the profiles of the DataVaults app and information gathered from ongoing Municipality of Piraeus projects, we aim to create novel services that target tourists and citizens visiting the city of Piraeus.

3.3 HEALTHCARE DATA RETENTION AND SHARING

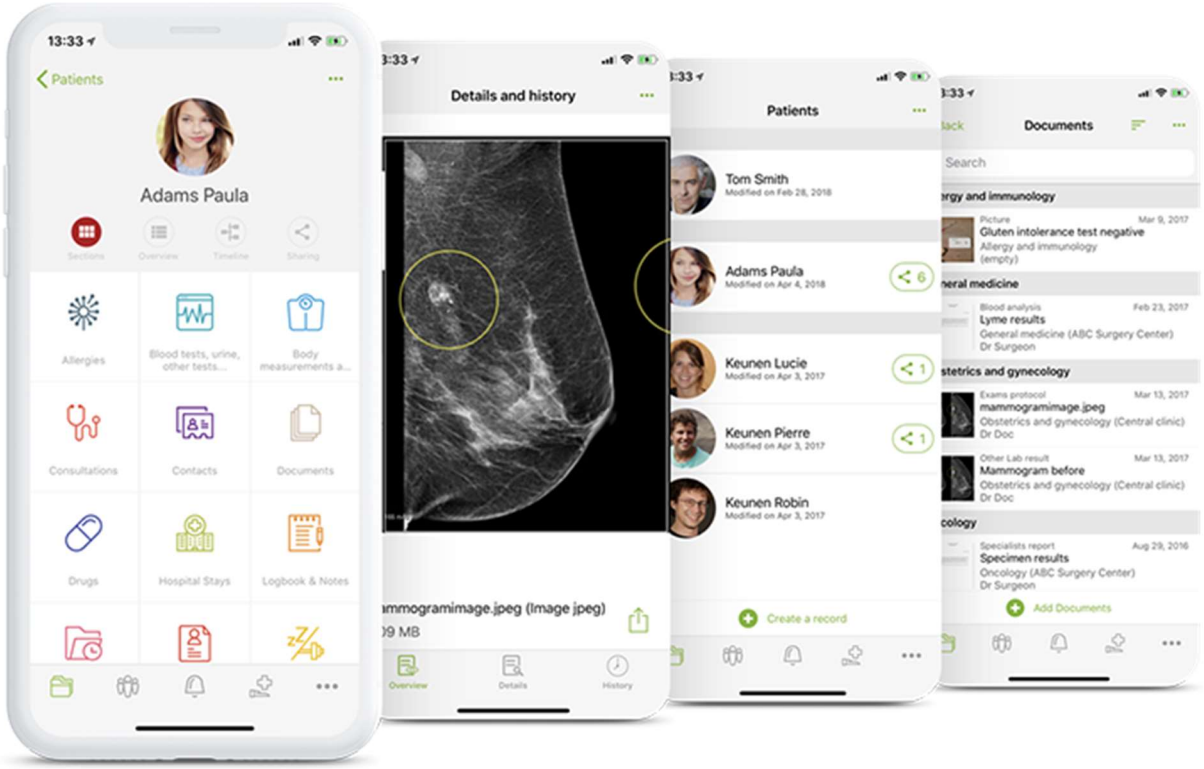

Figure 7 - Andaman7 application

PROBLEM STATEMENT

Healthcare data is very sensitive and personal. Andaman7 started by storing health data of patients on their smartphone only (server-storage less system). However, there is a need for patients to share data, both for care and for medical research: most patients agree to share their data if they are asked for permission and their data is not abused (some patients are open to share their data if they can get some profit).

With Andaman7, patients manage their complete health data - by that we mean medical data (allergies, medications, conditions, consultations, medical history, documents of all kinds, lab tests results, etc; in short, the content of the hospital EHR) but also fitness data (collected from other mobile apps and connected devices), sleep, nutrition etc.

The data model is schema-less, so very extensible by nature. Legal and ethical issues are also very important aspects, especially for data as the ones used in Andaman7.

CONSIDERED SCENARIO

The following scenarios will be pursued during the project:

- **Scenario a**. Current users of Andaman7 will be able to connect to DataVaults to store all or part of their health data (after explicit consent). 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).
- **Scenario b**. Current users of Andaman7 will be able to connect to DataVaults to collect their data (coming from various sources) and store them in Andaman7 on their smartphone. 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 (e.g. result of a clinical trial, comparison to a specific group, …)

DataVaults offering to demonstrators

By integrating Andaman7 and DataVaults, users will be able to make better use of their data (share data for care and/or research in a broad sense) and will contribute to improve health. It will also allow collecting additional data (aggregated or not), store them in Andaman7.

This enriched data will be usable for additional types of data treatments (AI based techniques, big data discovery, personal medicine, etc.)[1][2].

[1] https://ec.europa.eu/research/health/pdf/factsheets/real_world_data_factsheet.pdf

[2] https://www.researchgate.net/profile/Enrique_Bernal-Delgado/publication/305073029_Making_sense_of_big_data_in_health_research_Towards_an_EU_action_plan/links/5780b98408ae69ab88260249/Making-sense-of-big-data-in-health-research-Towards-an-EU-action-plan.pdf

3.4 SMARTHOME PERSONAL ENERGY DATA

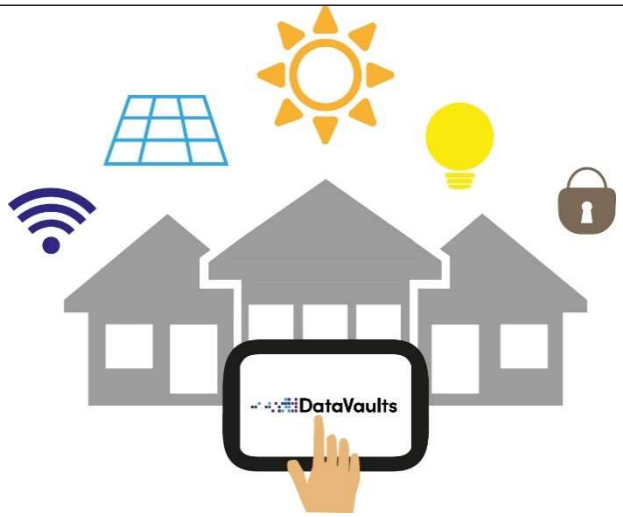

Figure 8 - DataVaults Smarthome

Scenario 1: PV installation design for self-consumption

PROBLEM STATEMENT

Nowadays, thanks to new regulation and the increasing awareness about environmental issues and climate change, photovoltaic (PV) systems are highly demanded. Producing your own energy with low environmental impact, decreasing the CO_2 emissions and the electricity bill at the same time is something very interesting for all type of consumers, residential, large industries or tertiary buildings [10]. Currently, in order to design an adequate PV installation our company needs not only the electricity consumption of the buildings but also some extra information: available space, kind of roof (slope or flat), location, etc. Then, if a customer or possible customer wants a PV installation, they have to contact us and fill a template to provide all this information.

Currently we use the consumption data just for billing purposes, to foresee the future energy demand and to make basic recommendations inside each client's intranet, like a change of tariff to have a lower price. The amount of energy data collected for each client is huge and varies depending the kind of tariff and the meter installed by the DSO.

Nowadays we use the consumption data just for billing purposes, to foresee the future energy demand and to make basic recommendation inside each client intranet. These data are not currently used to profile clients and try to get extra information for their daily activities. This extra information could help to understand better our clients and interact with them.

CONSIDERED SCENARIOS

- **Scenario 1: PV installation design for self-consumption:** The user can introduce all the required information through DataVaults in exchange of some bonus or compensation. With this information and the energy consumption, a preliminary but fairly accurate PV design can be made. A non-company client could also provide their energy consumption with the platform so we could perform the same study. The design will show to the user the capacity of the PV plant, the area needed to install it, and as a

result, the estimated production and the economic savings. Using ratios according to total energy consumption the capacity of the plant can be calculated. Comparing the estimated energy production of that photovoltaic system with the energy consumption each hour, energy and economic savings can be projected.

- **Scenario 2: Improve profiling of clients to enhance energy efficiency:** We could offer services with extra value to our clients if we analyse their consumption data along with useful additional information of the dwelling gathered through DataVaults platform such as area (m^2) or number of people. Clients profiles and segmentation can be done. It would be useful if the platform could provide automatically energy ratios (kWh/m^2 and kWh/person). These ratios will indicate an estimation for energy efficiency and will show which clients have potential for improvement. We could use these ratios to offer solutions to non-efficient users, but also external energy consultant companies could use these ratios to offer their services.
  Clients could have a ranking to classify their ratios, been able to check if they have a better efficiency index and act according it.
- **Scenario 3: Energy consumption patterns with personal preferences:** Analysing the hourly energy consumption routines of the users can be made for each period, having a model day, model week, etc. With additional information of personal data and preferences of the client like worktime, what they do in their spare time, reading, practising sports, food taste, etc… Service companies can adjust their products to match the user's ones. For example, if a restaurant has access to this combined information, they can make offers the days which the data providers usually go out for dinner, or offer a different type of food because they know people living in a specific neighbourhood have a healthy way of life. In this third scenario, extra smart-home devices such as smart plugs that measure specific loads, could be installed to have better understanding of the user's patterns (when the TV is on, how much time they spend cooking…)

SOLUTION

DataVaults gives our company the opportunity to implement a database with all this information and analyse the possibility of installing a PV plant. Another option could be giving this information, with the approval of the data provider (data owner), to other companies to offer these services.

With DataVaults we could get all the information required to design a PV plant faster and more accurate. Then we won't have to wait for the customer to ask for this service, hence, the company can have a more proactive attitude towards the DataVaults users, being able to approach the data providers with better ROI to offer the PV installation.

Another positive result for our company is promoting PV installation among the users with the most profitable outcome would give us credibility and an improve image in the energy market

The implementation of the second scenario can reduce the total amount of energy consumption, and the idea of doing a ranking like a competition will stimulate the participation of households in order to reduce the demand. The presence of clients inside the intranet will be higher cause of the interactive area, making them more active and enhancing

the communication between our company and our customers. These ratios can be sold to external energy consultant to focus their efforts in the worst efficiency users.

For the third scenario, consumption data get an extra value combined with personal one. Having access to it, not only our company but also other collaborating companies can offer custom service to the DataVaults users, that fit their way of life. At the same time, the routine of the clients can be integrated in our demand estimation software which we use to calculate the future energy that we have to buy in the electricity market and make it more accurate, saving money four our company.

3.5 PERSONAL DATA FOR MUNICIPAL SERVICES AND THE TOURISM INDUSTRY

PROBLEM STATEMENT

Currently the offices responsible for the digital services of the Municipality of Prato (such as tax payment, road ticket payment, issuing of certificates, school canteen registration, etc.) carry out customer satisfaction surveys by building sample lists of users, not only resident in Prato, to which questionnaires are sent. Users' identification data (telephone, email, etc.) is not always updated, while not all users provide their consent to be contacted for this type of investigation. The Municipality would therefore like to be able to facilitate the updating of users' contact data and widen the pool of those available to contribute to customer satisfaction surveys.

Moreover, the event organization of Cultural and tourist institutions in the city could be better coordinated, also due to the current impossibility of knowing how the users of these events (tourists, spectators, visitors) move around the city, participate in events, which are most appreciated by type and quality. The operators of these entities therefore find it difficult to improve their offer on the basis of a more in-depth knowledge of the target audience.

CONSIDERED SCENARIO

- **Scenario 1: Customer satisfaction analysis for the administration services:** Users can manage their own data and share it with the municipal registry service (in the case of a resident citizen), in order to check and update it if necessary. In this case, the updated data on DataVaults is automatically used also for the registry services, which interface with DataVaults. In the event that the user is not a resident in Prato (and therefore not present in the registry), through the app he/she can still enter his contact details and personal data.

  Alternatively, the citizen residing in Prato goes to a registry office where he/she updates his/her data: at the counter he/she is suggested to download the Prato app to help the municipality to keep the data updated in case of need. The data that is updated at the counter is automatically propagated to the DataVaults infrastructure through proper APIs, respecting the requirements of the data owner on its usage by the Municipality.

  By using the app daily, the user helps to provide data continuously in his/her DataVaults repository at least in the following ways:

1. Form for voluntary updating of personal data
2. Location information at intervals
3. Detection of the addresses corresponding to the mail accounts configured on the phone

The service operator who must carry out a customer satisfaction survey uses the web interface of his service to extract a large list of potential interviewees, submits the list to the DataVaults platform through the appropriate web interface, also composing the form with the questions to be administered. The users of the list will receive an invitation to participate in the survey and, in case of consent, will fill in the relative form.

The user included in the sample list receives in the app a notification of the proposal of a contract that regulates the use of his/her data by the Municipality, through which he/she will obtain a fair remuneration. Through the app, the user accepts the proposal and receives a survey questionnaire for customer satisfaction from the Municipality. Once the questionnaire is completed, the user receives the equivalent agreed in the contract in his/her digital wallet on DataVaults.

- **Case 2: Approval and use of cultural and tourist services in the city:** Cultural and tourist operators access the web interface of the DataVaults platform to set up news relating to the various cultural events they organize, scheduled in the city.

  The user installs the Prato app on his/her smartphone and with the daily use of the app he/she helps to provide data in at least the following ways:

  1. form for voluntary updating of personal data
  2. location information at intervals
  3. detection of the addresses corresponding to the mail accounts configured on the phone.

Figure 9 - Satisfaction analysis and tourist services in Prato

Cultural and tourist operators can use the platform in the following ways:

- Targeted information on scheduled events: the operator sends the users who meet certain selection criteria based on the data present on the DataVaults platform (e.g. position, personal data, movements in the city, etc.) information banners relating to the scheduled events;

- Request for feedback on attended events: always on the basis of the above selection criteria, the operator can request the expression of an opinion/approval for a given cultural event attended by the user;
- Data analysis: the operator can extract from the DataVaults platform analytical information relating to, for example, typical itineraries of tourists in the city, statistics on the attenders of an event, correlations in the attendance of events, satisfaction, etc.

Furthermore, through the interaction between the Prato app and social apps on the user's smartphone like Facebook or Twitter, the cultural/tourist operator can push some specific content on the attended event and invite him/her to share such content on his/her own social networks. This will increase publicity and knowledge on cultural events in the city.

Upon the request for use of his/her data, the user receives a notification of a smart contract on his/her smartphone, thanks to which he/she will receive appropriate remuneration for the activities voluntarily carried out. The amount defined in the contract is charged to the user's digital wallet on DataVaults.

SOLUTION

The implementation of the DataVaults approach will help the Municipality of Prato to better design its digital services, in order to meet citizens' expectations and to ensure a continuously updated flow of information between the administration and the service users, in a renovated framework of mutual trust and engagement.

It will also offer a valuable approach for revising the current methodology for managing the customer satisfaction procedure, in order to make it easier for the internal staff and a more participated one, thanks to the wider involvement of service users.

The DataVaults approach will increase the planning and management of cultural events in the city, due to the opportunity given to local cultural institutions to better investigate touristic flows and to collect a large set of comments regarding their offer.

It will also give tourists and citizens attending cultural events the chance of getting personally involved by providing feedbacks and appreciations also through their social channels, in a perfectly safe and privacy respectful context. This will reinforce the trust between cultural service providers and service users, the latter becoming the real centre of the process.

4 PERSONAL DATA VALUE CHAINS

In "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value" [6] the author identifies 3 different categories of personal data from the point of view of the gatherer platform. These 3 types of data are the following:

- **Volunteered or surrendered**: Data which is explicitly shared by the user about themselves. Some examples given in the document are: "when someone creates a social network profile, enters credit card information for online purchases, provides his/her personal information as a condition of registration to a given on-line service, or posts information about a friend, colleague, family member, etc"
- **Observed**: This category of data contains data that is gathered by recording the activities of users. Examples given are: "Internet browsing preferences, location data when using cellular mobile phones or telephone usage behaviour, etc"
- **Inferred**: Data that is extracted by analysing the data in the rest of the categories, sometimes even using other inferred data. Examples: "advertising profile, credit score, etc"

All these categories of data can be collected and will enter the data value chain. In the data environment there is a consensus on the high-level data value chain definition. Every document provides a similar approach to the definition. We will take as base, the data value chains defined by the BDVA, as shown in Figure 1. It is always similar as the diagram in Figure 10 that has been extended to show how the project fits that data value chain.

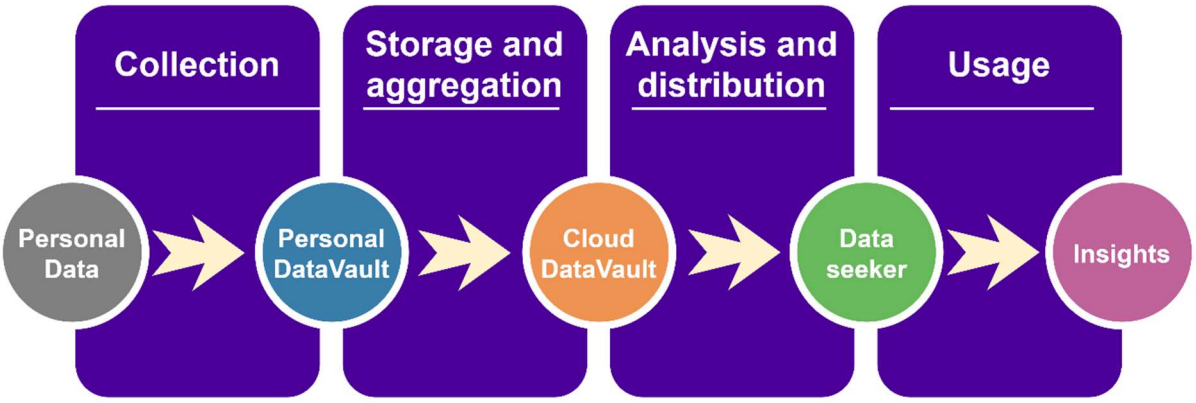

Figure 10 - General data value chain

Starting from the scenario's description, the possible pipelines to be implemented will be identified. Then, with the help of the demonstrators the idea is to identify which current data sources can be used and what extra data sources are needed to fulfil that pipeline.

In order to do so, we must look into the State-of-the-art about these data platforms to help us identify possible technologies and methodologies to implement those data pipelines. This process is shown as a diagram in Figure 11.

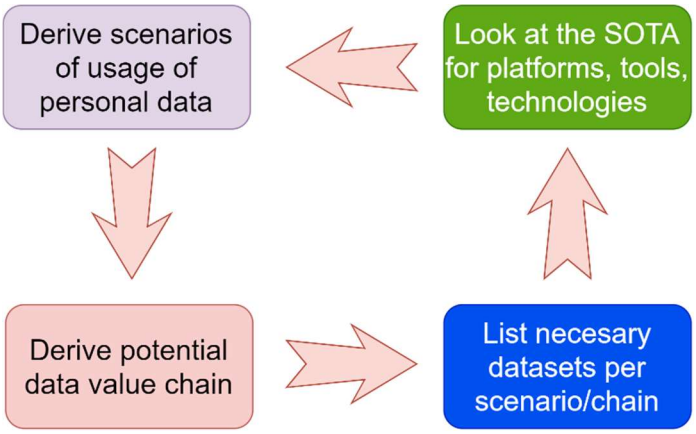

Figure 11 - Methodology to define the data value chain

Moreover, with the help of the demonstrators, we also need to identify and describe the possible data sources needed to fulfil the scenario. Not only the ones that are currently in use that can be moved to DataVaults, but also the ones that can help them to complete the use case described in the scenarios but are not yet used by them. These needed data sources could be some that the user volunteers within the scope of their personal data, but they can also be observed or inferred external data sources that will help the demonstrators to create a better value for the users, such as social media data.

4.1 DEMONSTRATORS PERSONAL DATA VALUE CHAINS

Considering the previously defined data value chain definition and focusing now in the demonstrator scenarios of this project, we will define a set of high-level data value chains and pipelines that compound the core functionalities to be expected from DataVaults.

During the course of the project these data value chains will be further detailed to contemplate extra functionalities extracted from the scenarios.

We have identified 2 use cases that are shared among all the scenarios:

1. **A user wants to use DataVaults to store their personal data and make it shareable at will**

   This is the main functionality of the DataVaults project from a user's point of view: A user is ready to use DataVaults and wants to store some or all of their data into DataVaults. Since the data to be stored will surely be heterogeneous and different users will try to share their data with different formats and structures, the first step of the data value chain is to homogenize that data in order to store it in the DataVaults Data Model as is modelled in point A of Figure 12.

   Once the data is in the expected format to fit our Data Model it is stored in the user's personal vault. This is a local storage where the data rests until the user decides if they want to share it with the cloud storage as modelled in point B. In order to upload the data to the secure cloud storage, the user is first asked for permission to do so. If this permission is not granted the data will stay in the personal storage until the user's opinion change. If the user decides to grant permission, the data is then uploaded to the secure cloud storage within DataVaults as can be seen in point C of Figure 12.

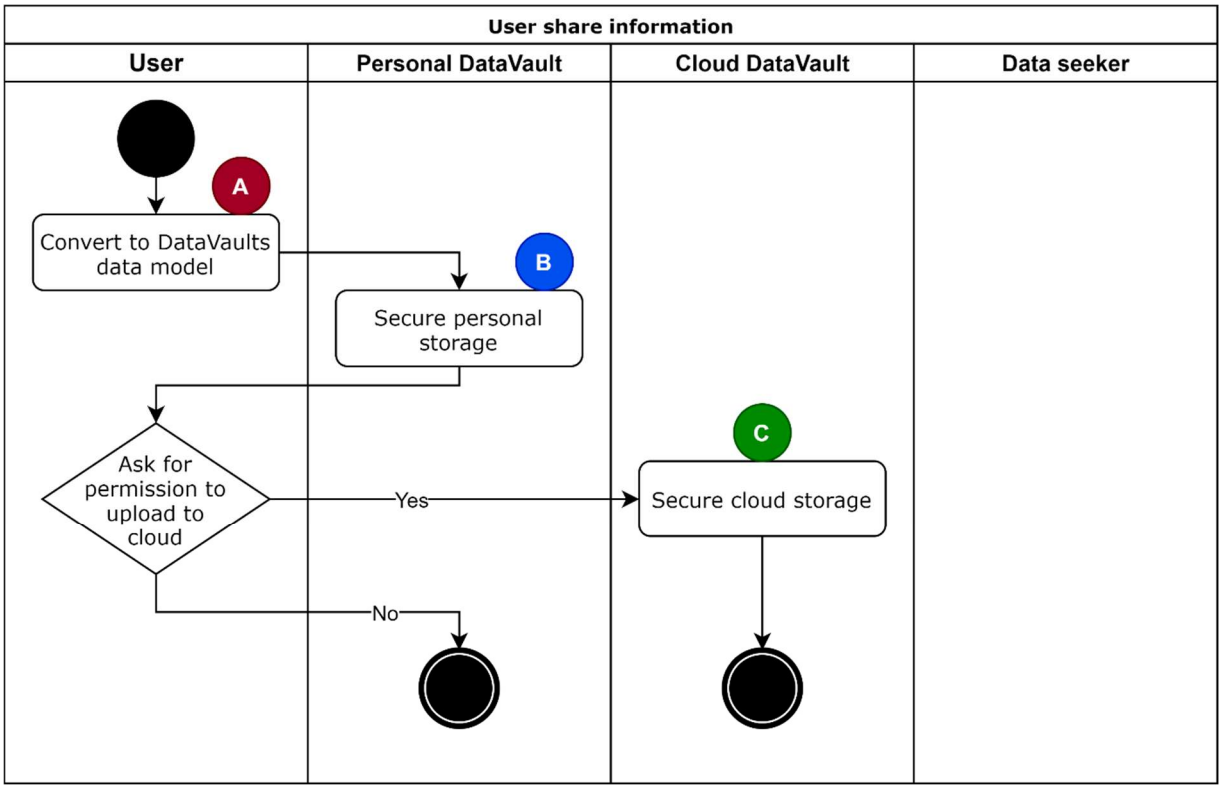

Figure 12 - User sharing Data Value Chain

2. Data seekers want to get data from DataVaults

This will be the main functionality from a data seeker point of view: requesting data to DataVaults to perform their analysis. In order to do so, the data seeker might consult the metadata available in the DataVaults platform. Since the project foresees 2 separated Vaults, one on the cloud and another one on the user side, the platform will need to look into both to answer the data seeker request as can be seen in point C of Figure 13.

The next step of the chain would be to ask the user for permission to use the data. In order for the data seekers to gather the data, they need the user's permission to do so. This permission could be pre-granted in advance by the user to certain uses. This permission logic is represented in the Figure 13 in Point B.

Once the consent is given by the user, it will be stored in the platform. In parallel, the platform will start working on the data processing requested by the data seeker before the data is sent to them. The DataVaults platform will have some analytics and pre-processing capabilities that the data seekers will be able to use. Once the requested processing has been done by the platform, the data will be sent to the data seeker. This part of the chain is represented in point C of Figure 13.

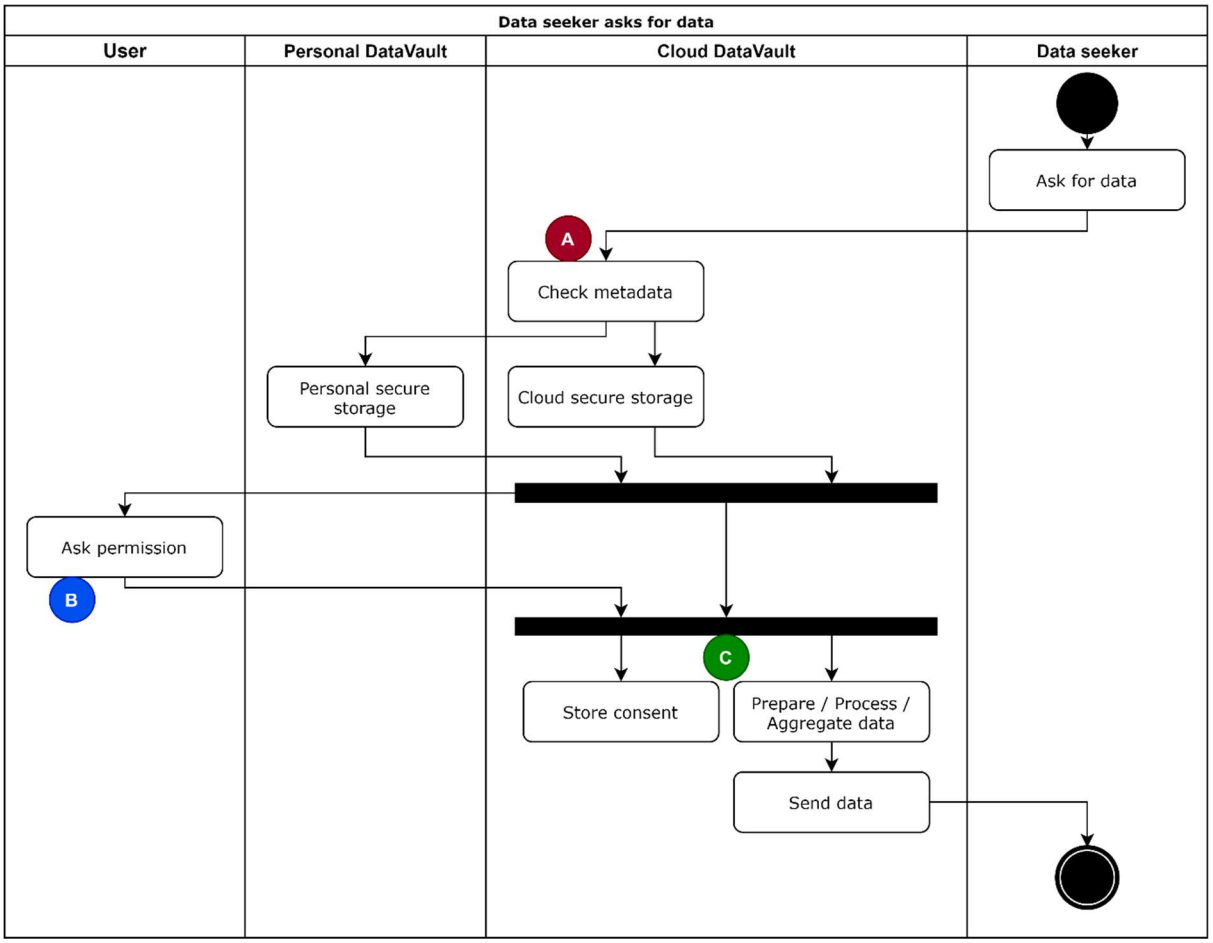

Figure 13 - Data seeker Data Value Chain

As explained, the previous diagrams represent a holistic version of the data value chains: the first one from the point of view of a user wanting to store user in the DataVaults platform, the second one from the point of view of a data seeker wanting to get some data from the platform.

In the Figure 14 we integrated both detailed versions in a consolidated version that summarizes the core data value chain that will form the DataVaults project. In this diagram we have integrated the concept of compensation to the user. Every time a data seeker asks the platform for data, they need to set a compensation to be given back to the users whose data is used. This way, the project will set up a sort of data market where the user can decide if the compensation to be received from the usage of their data is enough for them so they can decide, with enough knowledge, whether to give permission to the data seeker or not.

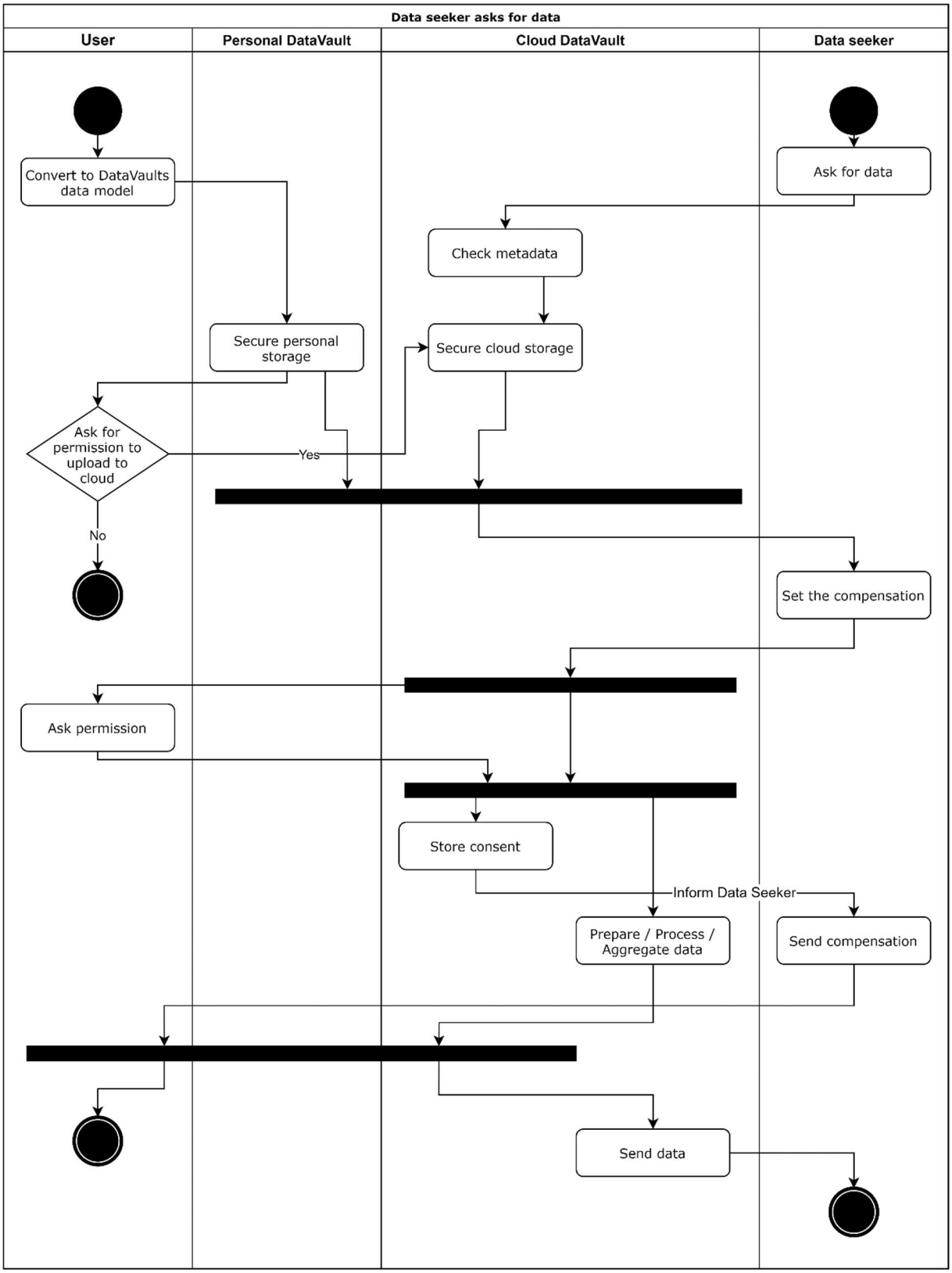

Figure 14 - Core DataVaults data value chain

4.2 INITIAL DEMONSTRATOR REQUIREMENTS

The gathering of the initial demonstrator requirements involves the re-working of the basic scenarios set out in Section 3 above, in order to establish the mechanism and forms of communication to be able to offer assistance in the software development process. This

specific section is looking to provide a textual description of the demonstrator requirements for the data platform in order to realise the scenarios and data value chains described above.

The requirements from the demonstrator's perspective are at the core of the project. They emerge from T1.3 and morph into T5.1 which is devoted to systematically aggregate and analyse the user requirements of all stakeholders involved in the DataVaults value chain, so as to set out what is needed in the DataVaults platform.

The methodology used in order to gather the requirements of the platform from the demonstrators' perspective was determined by the adoption of an "agile software development approach". This necessitates having a mechanism for dialogue between the demonstration sites and the technical teams which will be ongoing throughout the project. Investment of time at the beginning of the project to closely identify with the scope of the project and to create the initial prioritized stack of requirements is imperative.

The process of gathering the initial requirements, started in this WP, will form the basis for further elaborations as the project matures, with solutions evolving through the collaborative effort and ongoing dialogue of all partners. Such an approach advocates adaptive planning, evolutionary development, early delivery, and continual improvement and it encourages rapid and flexible response to change. The high levels of iteration, collaboration and flexibility should reduce risks associated with requirement capture.

4.2.1 User stories

The starting point for this dialogue will be the user stories provided by each of the demonstration sites, assisted by their technical partners, which can be found in Table 2 below. As the project proceeds, iterations will occur and the stories will evolve, whilst other techniques will be added to enhance the stories, such as "user interviews", "questionnaires" and "story writing workshops" for example. This will make them more valuable as a means of communication between the pilot sites and those developing the technology.

The reported usage scenarios and use cases will facilitate the elicitation of the functional and non-functional requirements of DataVaults, including the requirements of the platform from the demonstrators' perspective, taking into account the views of all stakeholders, At this stage of the project, we are only concerned with gathering a relatively high-level view of the demonstrator's requirements in order to start the conversation process. But the user stories which we gather for this D.1.1 deliverable, will remain in focus throughout the project, and they will grow and develop with finer detail at every iteration to ensure the best options are available for the demonstrations. These iterations continually assist the project managers in quickly being able to understand what is expected by the demonstrators regarding the various elements of DataVaults.

User stories are brief and short, and the custom was to have them written up on small cards or post-it notes to start the process. They are not meant to be lengthy or contain descriptive text and the focus is upon "Who? What? & Why?" of a feature in order to give direction to the developing teams of how they should proceed. They are essentially short descriptions of functionality told from the user's perspective. The focus is on *why* and *how* the user interacts

with the software. A user story is essentially a high-level definition of what the software should be capable of doing. They usually follow a template similar to that we have adopted which is:

As a <type of user>, I want <some desired outcome> so that <some reason>.

Essentially, a user story allows emphasis to move from capturing requirements on paper to discussing in business language, the required functionality.

The demonstrators will be providing both their functional requirements and non-functional requirements. A non-functional requirement being one that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors of it which is the case with functional requirements. Some typical non-functional requirements are:

- Performance Scalability.
- Capacity.
- Availability.
- Reliability.
- Recoverability.
- Maintainability.
- Serviceability.

Together they ensure the usability and effectiveness of the entire system. Iterations of the "user stories" below will subsequently tease out the greater detail required.

4.2.2 User Stories provided by the demonstrators for identification of technical requirements

OLYMPIACOS: Sports and activity personal data		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a fan	*I want to store a copy of my personal data on a highly secured cloud storage*	*To back it up.*
As a fan	*I want to control my personal data and social media activity*	*To control who can have access to a specified portion of my data.*
As a fan	*I want to preconfigure data seeker categories*	*To control who can have access to a specified portion of my personal data.*
As a fan	*I want to control my data*	*To determine what part of my data can be shared to third parties.*
As a fan	*I want to keep track of all data transactions*	*So that that I give consent to ensure security and trust.*
As a fan	*I want my data to be fully anonymized or pseudonymized before sharing to third parties*	*To protect data being identified as being mine.*
As a fan	*I want to download on my smartphone all personal data uploaded by various sources*	*To securely store and keep all aspects my personal record up to date.*

	on a highly secured cloud storage	
As a fan	*I want to have app functionalities on my smartphone*	*So that I can check and update my personal data in the administration CRM system*
As a fan	*I want to receive a notification on my smartphone when my personal data is accessed*	*So that I am aware of whom is using it and why*
As a fan	*I want to agree on the usage of my personal data*	*So that I can get rewarded by those using it*
As a fan	*I want to have full control over my personal data accessibility*	*So that I can decide and authorise the level of disclosure it should have*
As sport club	*I want to know about my customer´s likes*	*So that I can improve my services to them*
As sport club	*I want to know about my customer´s habits and interests*	*So that I can implement the most suitable marketing campaigns, improve the fan experience and build a stronger relationship with them*
As sport club	*I want to access anonymised personal data*	*So that I can build statistics over it*
As sport club	*I want to access updated contact data of service users*	*So that I can inform the about the match schedule and the Club General Assemblies*
As sport club	*I want to access different rewarding options*	*So that I can propose the most suitable to the fans*
As sport club	*I want to know the level of fan satisfaction through surveys*	*So that I can examine the level of the fan satisfaction and improve the provided services*
As sport club	*I want to access anonymised personal data*	*So that I can approach the interest of new sponsors*
As a data seeker (sponsor)	*I want to access anonymised personal data*	*So that I can organize my strategies better based on the customers' profiles*
As a data seeker (sponsor)	*I want to access fan personal data of service users*	*So that I can send them advertisement about the sponsor products/services*
As a data seeker (sponsor)	*I want to access fan location and social media data*	*So that I can organize and run events and campaigns focused on the fans' interests*
As a data seeker (sponsor)	*I want to analyse fan data*	*So that I can build user profiling for market needs*
Scenario B. Athletes		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As an athlete	*I want to store a copy of my personal data on a highly secured cloud storage*	*To back it up.*

As an athlete	*I want to control my personal data*	*I want to control who can have access to a specified portion of my data.*
As an athlete	*I want to control my personal data*	*I want to approach the interest of new personal sponsors*
As an athlete	*I want to preconfigure data seeker categories*	*To control who can have access to a specified portion of my personal data.*
As an athlete	*I want to control my data*	*To determine what part of my data can be shared to third parties.*
As an athlete	*I want to keep track of all data transactions*	*So that that I give consent to ensure security and trust.*
As an athlete	*I want my data to be fully anonymized or pseudonymized before sharing to third parties.*	*To protect data being identified as being mine.*
As a coach	*I want to access athletes personal and sport data*	*So that I can plan the training sections and the tactics based on the athletes profiles*
As a coach	*I want to access athletes personal and sport data*	*So that I can scout athletes to improve the quality of the squad*
As a data seeker	*I want to access anonymized sport data of multiple athletes*	*To help improve the sport sector.*
As a data seeker	*I want to access analytics sport data of multiple athletes*	*To help improve the sport sector.*
As a health data source (e.g. hospital)	*I want to add some health data to an athlete's health record.*	*To ensure that it is complete and up to date.*
***PIRAEUS:* STRENGTHENING ENTREPRENEURSHIP AND MOBILITY**		
Scenario A. Smart Mobility Services for Individuals		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a citizen- fan	*I want to have app functionalities on my smartphone*	*So that I can avoid traffic during events*
As a citizen- fan	*I want to have app functionalities on my smartphone*	*So that I can be notified on events happening near the stadiums*
As a citizen- fan	*I want to have app functionalities on my smartphone*	*So that I can provide personal data in addition to that included by law in the population registry*
As a data seeker (Municipality or third party)	*I want to access anonymized traffic and route data of multiple citizens-fans, during events*	*To study flows during events and propose decongestion solutions*
As a local government institution	*I want to have a mobility plan during events*	*To better serve citizens-fans*

As a sports club	*I want to provide information to the fans*	*To avoid last minute congestions and enhance the fan experience*
Piraeus Scenario B. Empowering local entrepreneurship		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a citizen	*I want to have app functionalities on my smartphone*	*So that I can be notified on sale deals*
As a citizen	*I want to have app functionalities on my smartphone*	*So that I can search nearest shop location*
As a data seeker (Municipality, Local Traders Association, Local Chamber of Commerce)	*I want to access anonymized data of multiple citizens regarding shopping preferences*	*To analyze the data and produce indicators on shopping trends and behaviors*
As a data seeker (Municipality, Local Traders Association, Local Chamber of Commerce)	*I want to access anonymized data of multiple citizens regarding shopping preferences as well as derivatives of these data (see case above)*	*To organize events, sales and promote the local trade center*
As a data seeker (Individual traders)	*I want to access anonymized data of multiple citizens regarding shopping trends*	*To provide instant sale opportunities and personalized*
As a local government institution	*I want to provide information and tools to the local stakeholders*	*To enhance local entrepreneurship and local economy*
Piraeus Scenario C. Services for personalized cultural and touristic experiences		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a citizen/tourist	*I want to have app functionalities on my smartphone*	*So that I can be informed on events and sights based on personal cultural and recreational interests*
As a citizen/tourist	*I want to have app functionalities on my smartphone*	*So that I can communicate whether I attended an event or share my tourist experience*
As a citizen/tourist	*I want to have app functionalities on my smartphone*	*So that I can communicate whether I liked or not an attended event or an experience*
As a citizen/tourist	*I want to have app functionalities on my smartphone*	*So that I can plan a sightseeing route (walk, bus or alternate)*
As a destination management organization	*I want to access anonymized personal data of tourists*	*So that I can build statistics over It (entertainment*

		preferences, shopping behavior, places visited, likes and dislikes etc.)
As a destination management organization	*I want to create derived indicators and statistics combined with anonymized personal data of tourists*	*So that I can plan a destination management strategy based on existing or upcoming touristic flows*
As a destination management organization	*I want to access citizen/tourist personal data of service users*	*So that I can communicate scheduled cultural events based on user preferences*
As a destination management organization	*I want to push a notification on user smartphone*	*So that I can get feedback on my promotion strategy*
Piraeus: All scenarios		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a user	*I want to control my data*	*To control who can have access to all or portion of my data.*
As a user	*I want to preconfigure data seeker categories*	*To control who can have access to all or portion of my data.*
As a user	*I want to control my data*	*To determine what part of my data can be shared to third parties.*
As a user	*I want to keep track of all data transactions*	*I give consent to ensure security and trust.*
As a user	*I want my data to be fully anonymized or pseudonymized before sharing to third parties.*	*To protect data being identified as being mine.*
As a user	*I want to be notified before consenting to share data, in case where combination with pre-shared data can lead to identification.*	*To protect from identification.*
MIWENERGIA: **SMARTHOME PERSONAL ENERGY DATA**		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a customer	*I want to know information about my energy consumptions*	*So that I could modify my consumptions times and spend less money*
As a customer	*I want to have access to my energy bills*	*So that I can check what I am spending on*
As a customer	*I want to have recommendations to save energy*	*So that I can save money*
As a customer	*I want to have information about renewables systems*	*So that I can spend less energy and help the environment*

As a customer	*I want to know my energy efficiency*	*So that I can improve my energy efficiency*
As a customer	*I want personals recommendations according to what I like*	*So that I can find easily what I'm looking for*
As an electricity retailer	*I want to know about my customer´s likes*	*So that I can improve my services to them*
As an electricity retailer	*I want to know information about other electricity retailers' clients*	*So that I can offer them a better tariff*
As an electricity retailer	*I want to offer photovoltaic systems to my customers*	*So that I can improve the energy mix becoming greener and more attractive for the customers*
As an electricity retailer	*I want to know about my customer´s habits*	*So that I can estimate my global electricity demand better*
***ANDAMAN7:* HEALTHCARE DATA RETENTION AND SHARING**		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a user	*I want to store a copy of my health data on a highly secured cloud storage*	*To back it up.*
As a data seeker	*I want to access anonymized health data of multiple patients*	*To help improve the health sector.*
As a data seeker	*I want to access analytics health data of multiple patients*	*To help improve the health sector.*
As a user	*I want to control my health data*	*I want to control who can have access to a specified portion of my health data.*
As a user	*I want to preconfigure data seeker categories*	*To control who can have access to a specified portion of my health data.*
As a user	*I want to control my health data*	*To determine what part of my health data can be shared to third parties.*
As a user	*I want to keep track of all data transactions*	*So that that I give consent to ensure security and trust.*
As a user	*I want my data to be fully anonymized or pseudonymized before sharing to third parties.*	*To protect data being identified as being mine.*
As a user	*I want to download on my smartphone all personal health data uploaded by various sources on a highly secured cloud storage*	*To securely store and keep all aspects my personal health record up to date.*

As a user	*I want to control my personal online health record.*	*To determine who can add health data to it*
As a user	*I want to keep track of my personal health record.*	*To monitor all the changes (what, who and when) done to it.*
As a health data source (e.g. hospital)	*I want to add some health data to a patient health record.*	*To ensure that it is complete and up to date.*
As a user	*I want to know what other people correspond to my profile.*	*In order to compare my results to those with a similar profile to me.*
As a user	*I want to get statistical results from other people corresponding to my health profile.*	*To compare my results with them.*
PRATO: MUNICIPAL SERVICES AND THE TOURISM INDUSTRY		
Scenario 1 - Customer satisfaction analysis for the administration services		
Subject/Who?: "As a <type of user>"	***Action/What?: "I want <some desired outcome>"***	***Reason/Why?: "so that <some reason>"***
As a citizen	*I want to have app functionalities on my smartphone*	*So that I can check and update my contact data in the population registry*
As a citizen	*I want to have app functionalities on my smartphone*	*So that I can check and update my personal data in the administration CRM system*
As a citizen	*I want to have app functionalities on my smartphone*	*So that I can provide personal data in addition to that included by law in the population registry*
As a citizen	*I want to visualise all my personal data on my smartphone*	*So that I can understand who owns and manages it*
As a citizen	*I want to visualise a map of all my personal data on my PC*	*So that I can understand who owns and manages it*
As a citizen	*I want to receive a notification on my smartphone when my personal data is accessed*	*So that I am aware of whom is using it and why*
As a citizen	*I want to select the type and frequency of notifications on the usage of my data*	*So that I can limit those to the most important ones*
As a citizen	*I want to agree on the usage of my personal data*	*So that I can get rewarded by those using it*

As a citizen	*I want to have full control over my personal data accessibility*	*So that I can decide and authorise the level of disclosure it should have*
As a citizen	*I want to access my Personal Wallet*	*So that I can check my rewards for the use of my personal data*
As a citizen	*I want to perform the same actions both through the app and on my PC*	*So that I feel more comfortable*
As an administrator of digital services in the municipality	*I want to access anonymised personal data*	*So that I can build statistics over it*
As an administrator of digital services in the municipality	*I want to access updated contact data of service users*	*So that I can send them a survey for customer satisfaction*
As an administrator of digital services in the municipality	*I want to access different rewarding options*	*So that I can propose the most suitable to the users*
As an administrator of digital services in the municipality	*I want to download the results of the survey*	*So that I can store it for further use*
***Prato: Scenario 2* - Approval and use of cultural and tourist services in the city**		
As a citizen/tourist	*I want to have app functionalities on my smartphone*	*So that I can provide personal data such as contact data and cultural interests*
As a citizen/tourist	*I want to get a prompt on my smartphone*	*So that I can communicate whether I attended a given event or not*
As a citizen/tourist	*I want to get a prompt on my smartphone*	*So that I can communicate whether I liked or not an attended event*
As a citizen/tourist	*I want to get a connection with my social accounts*	*So that I can post comments over an attended event*
As a citizen	*I want to visualise all my personal data on my smartphone*	*So that I can understand who owns and manages it*
As a citizen	*I want to visualise a map of all my personal data on my PC*	*So that I can understand who owns and manages it*
As a citizen	*I want to receive a notification on my smartphone when my personal data is accessed*	*So that I am aware of whom is using it and why*
As a citizen	*I want to select the type and frequency of notifications on the usage of my data*	*So that I can limit those to the most important ones*

As a citizen	*I want to agree on the usage of my personal data*	*So that I can get rewarded by those using it*
As a citizen	*I want to have full control over my personal data accessibility*	*So that I can decide and authorise the level of disclosure it should have*
As a citizen	*I want to access my Personal Wallet*	*So that I can check my rewards for the use of my personal data*
As a cultural institution	*I want to access anonymised personal data*	*So that I can build statistics over it*
As a cultural institution	*I want to access citizen/tourist personal data of service users*	*So that I can send them advertisement on scheduled cultural events*
As a cultural institution	*I want to access different rewarding options*	*So that I can propose the most suitable to the users*
As a cultural institution	*I want to push a notification on user smartphone*	*So that I can get feedback on attendance to my cultural events*
As a cultural institution	*I want to access user location data*	*So that I can track cultural itineraries in the city*
As a cultural institution	*I want to issue a prompt on the smartphone of users attending my cultural events*	*So that I can suggest them to post some comment over the attended event on their social channels*
As a cultural institution	*I want to analyse user data per attended event*	*So that I can build user profiling for market needs*

Table 2 - User stories provided by the Demonstrators

5 DATA SOURCES

In this section we aim to list, for each demonstrator, the data sources needed to fulfil the scenarios and the data value chains presented in the document. This will be a high-level description, for now, with the main data sources identified by the demonstrators.

It is expected to work as the base information to start working on the drill down to more detailed information of those data sources, as well as the road to the detailed description of data pipelines and requirements for the scenarios.

Throughout the progress of the work done in this WP the list of current data sources will be detailed and formatted to suit the needs of the projects. Moreover, the list of needed data sources will probably be incremented and detailed depending on the picked scenarios, the data value chain and the requirements definitions.

This section summarizes the different data sources identified by the demonstrators and it is shown in a set of tables structured in the following way:

- Source: Defines what kind of data is gathered from this source, whether it is personal data, medical data, location…
- Type of data: Based on the data types defined in the BDVA reference model, as can be seen in yellow in Figure 1, they categorize the data into 6 groups:
  i. Structured data
  ii. Time series data
  iii. Geospatial data
  iv. Media, Image, Video and Audio data
  v. Text data, including Natural Language Processing data and Genomics representations
  vi. Graph data, Network/Web data and Metadata
- Volume: Estimation of the volumetric of the data source (Specified in records/ data size…)
- Origin: Where the data is coming from or from where is it extracted.
- Scenario: Identifies the scenario, from the ones described in Section 3, that makes use of the data source.

5.1.1 Sports and activity personal data

AVAILABLE DATA

Source	Type of data	Description of the data	Volume	Origin	Scenario
User/Personal data	Structured data	-Personal Data Information -Personal Governmental Data -Demographics -Contact details (Email, Mobile phone)	200000 records	CRM system (Microsoft Dynamics)	1

		-Personal Social Media Accounts -Member/Fan Number			
Data coming from medical exams	Structured data	Ergometric and medical exams -Statistical reports -Sport exercise/activity	2000 records		2

Table 3 - Olympiacos Available data

NEEDED DATA

Source	Type of data	Description of the data	Scenario
Social network	Structured data / Media	-User preferences -Activity and mobility data	1
Location data	Geospatial / Temporal	User position	1
Data coming from medical exams	Structured data / Media / IoT	Specialist report, tests, analysis	2

Table 4 - Olympiacos Needed data

5.1.2 Strengthening entrepreneurship and mobility

AVAILABLE DATA

Source	Type of data	Description of the data	Volume	Origin	Scenario
Population registry	Structured data	Demographics	400.000 citizens	Hellenic Statistical Authority	2,3
Personal Data	Structured data	-Personal Data Information -Profile -Contact details (Email, Mobile phone) -Personal Social Media Accounts	1.000 records	Pireapp	2,3
Location Data	Geospatial / Temporal	Live location of municipal buses	10 buses	Piraeus Intelligent City	1
WiFi access	Structured data	-Login Data -General location data	200 Access Points	WiFi access points	2,3
Feedback	General text	-Feedback -Opinions -Suggestions	100 Answers	All apps, social media and websites	1,2,3

Table 5 - Piraeus Available data

NEEDED DATA

Source	Type of data	Description of the data	Scenario
Personal data	Structured data	Any Personal data provided by the user: -Profile -Places visited -Buying preferences -Entertainment preferences	2,3
Location data	Geospatial / Temporal	Location Data through GPS or mobile phone to track live mobility through big events	1,2,3
Social network	Structured data / Media	-User preferences -User profile -Activity and mobility data	2,3

Table 6 - Piraeus Needed data

5.1.3 Healthcare data retention and sharing

AVAILABLE DATA

Source	Type of data	Description of the data	Volume	Origin	Scenario
User data	Structured data	Any health data (Specialist report, tests, analysis, consultation report, …)	UNKNOWN	Mobile app	1,2
Wearable data	IoT	Steps, sport activity, cardiac activity, ...	UNKNOWN	Mobile app	1,2
Data coming from health professionals / institutions	Structured data / Media	Specialist report, tests, analysis, consultation report, ...	UNKNOWN	Mobile app	1,2

Table 7 - Andaman7 Available data

NEEDED DATA

Source	Type of data	Description of the data	Scenario
Data coming from health professionals / institutions (new sources)	Structured data / Media	Specialist report, tests, analysis, consultation report, ...	1

Table 8 - Andaman7 Needed data

5.1.4 Smarthome personal energy data

AVAILABLE DATA

Source	Type of data	Description of the data	Volume	Origin	Scenario
Energy consumption data	Structured data	Consumption given by the DSO. Only our clients.	Unknown	DSO website.	All

Personal data	Structured data	Full name, bank account, address. Only our clients.	Unknown	The user by signing the contract.	All

Table 9 - MIWEnergia Available data

NEEDED DATA

Source	Type of data	Description of the data	Scenario
User data	Structured data	- Number of people per building - Size of the apartment - Worktime, preferences.	2 & 3
Dwelling	Structured data	- Available outside space - kind of roof (slope or flat). - Daily habits.	1

Table 10 - MIWEnergia Needed data

5.1.5 Personal data for municipal services and the tourism industry

AVAILABLE DATA

Source	Type of data	Description of the data	Volume	Origin	Scenario
Population registry	Structured data	- Personal data (name, address, fiscal code, place of birth, etc.) - Family data (members, relationships, etc.) - Citizens living abroad - Citizens no longer living in the city	250.000 citizens 200 GB	Database used by the municipality of Prato, to manage the population registry	1
Service contact data	Structured data	- Name/surname - Email - Mobile number - Contact channel choice for City Council services	61.500 Citizens 8,5 GB	CRM and eService authentication system of City Council	1
Social network	Structured data / media	- Posts - Contacts	500.000 persons	- Facebook - Twitter - Instagram	2
WiFi access	Structured data	-Hot spot Place -Number of connections in different time slots and places by user	35.000 users 1.000.000 Access-place 1 GB	Database managed by the Municipality of Prato	2

Table 11 - Prato Available data

NEEDED DATA

Source	Type of data	Description of the data	Scenario
Personal data	Structured data	- Any Personal data provided by the user (name, address, phone, email, etc) - Personal interests Email accounts stored on the smartphone	Both
Location data	Geospatial / temporal	User position	Both
Event description	Structured data	- Name of the event - Location of the event Day and Time of the event	1
Event participation	Structured data / Text	Text string / Boolean	1
Event evaluation	Structured data	Likert scale (1-5)	1

Table 12 - Prato Needed data

6 CONCLUSION AND FUTURE WORK

This document presents an overview of what a data value chain is, being more focused on the value of the personal data. As such, this deliverable starts with introducing the vision that the Big Data Value Association and the NIST Big Data Reference Architecture have about the value of the data as a first step to understand how the value of the data can be regarded within DataVaults, not only focused on the individuals but including the whole specter of the data. Both associations have common views on the phases adopted, such as data acquisition, store, analysis and usage.

As such, one of the objectives of the document is to present the state of the art of the data value chain. As a first approach to achieve this purpose, D1.1 is navigating around the different personal data platforms stakeholders, the personal data management platforms and the personal data methodologies and economy.

One of the most important ideas drawn from reading the state of the art of the data value chain, and more specifically of the value of personal data, is that there is no common defined methodology for calculating the value of personal data.

Individuals are also interested in knowing how much their personal data cost. In present, there are a lot of companies that obtain a high economic profit from our personal data without any remuneration to us for sharing these data.

The concept of privacy is also perceived in the document as one of the most relevant ideas. The individuals are open to paying more for better data privacy and security. The legal issues around the personal data will be one of the challenges of the project, some ideas have been reflected in the document and will be one of the discussion points in the future.

DataVaults aims for individuals to take control of their personal data, decide which kind of data can share, when and also obtain a reward for sharing it. As previously mentioned, the first step of DataVaults, as a platform focused on personal data, is to collect and store data coming from the different sources. 5 demonstrators are involved in the DataVaults project, Olympiacos, Piraeus, Andaman7, Miwenergia and Prato, with different interests around the personal data aspects, such as sports, mobility, healthcare, energy and the industry and thus will make use of diverse and different data sources.

DataVaults aims to establish a way to Improve the data management and handling of the different demonstrators, tackling the collection, formalization, storage, sharing and access control of these data in order to obtain a value. This value (or even profit) may be economic in some cases but also can improve other different aspects of society, for example using analytics algorithms to understand the way of how the people move around the city.

As we are at the beginning of the project, more details on the technologies that DataVaults is going to use has to be provided once the initial requirements of the demonstrators are set and an MVP for the overall platform is defined. This point will be one of the most relevant assets that the project has to achieve in the next months and will be driven by the other tasks of WP1, alongside with the technical implementation WPs that following.

7 REFERENCES

[1] M. E. Porter, Competitive Advantage: Creating and Sustaining Superior Performance, 1998.

[2] BDVA, "European Big Data Value Strategic Research and Innovation Agenda," 2017.

[3] NIST Big Data Public Working Group, "NIST Special Publication 1500-6 : NIST Big Data Interoperability Framework - Reference Architecture," 2015.

[4] I. Ng, "https://medium.com/hub-of-all-things/personal-data-as-an-asset-class-2e713deedd10," 1 11 2017. [Online]. Available: https://medium.com/hub-of-all-things/personal-data-as-an-asset-class-2e713deedd10. [Accessed 11 03 2020].

[5] World Economic Forum, "https://www.weforum.org," 11 02 2011. [Online]. Available: https://www.weforum.org/reports/personal-data-emergence-new-asset-class. [Accessed 11 03 2020].

[6] OECD Digital Economy Papers, "Exploring the Economics of Personal," *OECD,* 2014.

[7] A. Acquisti, "https://www.oecd.org/," 2010. [Online]. Available: https://www.oecd.org/sti/ieconomy/46968784.pdf . [Accessed 2020].

[8] UN, "Frontier Technology Quarterly: Data Economy: Radical transformation or dystopia?," 31 January 2019. [Online]. Available: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf.

[9] S.-A. Elvy, "PAYING FOR PRIVACY AND THE PERSONAL DATA," *Columbia Law Review.*

[10] UNEF, "http://www.unef.es," 09 2019. [Online]. Available: https://unef.es/wp-content/uploads/dlm_uploads/2019/09/memoria_unef_2019-web.pdf . [Accessed 12 03 2020].

[11] H. W. J. Cheng, M. LaFleur and H. Rashid, "Data Economy: Radical transformation or dystopia?," *Frontier Technology Quarterly,* no. 1, pp. 1-6, 2019.

[12] A. Acquisti, C. Taylor and L. Wagman, "The economics of privacy," *Journal of Economic Literature,* vol. 54, no. 2, pp. 442-492, 2016.

[13] T. K. M. P. A. P. Laura Kemppainen, "Emerging Revenue Models for Personal Data Platform Operators: When Individuals are in Control of Their Data," *Journal of Business Models,* vol. 3, no. 6, pp. 79-105, 2018.

[14] K. S. S. M. G. M. L. a. Y. G. N. Binh Truong, "GDPR-Compliant Personal Data Management: A Blockchain-based Solution," *IEEE TRANSACTION ON INFORMATION FORENSICS AND SECURITY,* 10 2019.