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

D6.6
DataVaults Scaleup Roadmap and Key Takeaways

Editor(s)	Sotiris Athanassopoulos, Nikos Achilleopoulos
Lead Beneficiary	MAGGIOLI
Status	Final
Version	1.00
Due Date	30/04/2023
Delivery Date	22/05/2023
Dissemination Level	PU

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

Project	DataVaults – 871755
Work Package	WP6 - Multi-Layer Demonstrators Setup, Operation and Business Value Exploration
Deliverable	D6.6 - DataVaults Scaleup Roadmap and Key Takeaways
Editor(s)	MAGGIOLI - Sotiris Athanassopoulos, Nikos Achilleopoulos
Contributor(s)	Suite5 – Galanos Vasilis Piraeus -Michail Bourmpos Prato – Paolo Boscolo Prato – Elena Palmisano MIWenergia - Borja Molina Rios MIWenergia – Ana Garcia Garre Andaman7 - Vincent Keunen Andaman7 - Sebastien Hannay Olympiacos - Christina Tsiligkiri Piraeus -Michail Bourmpos ETA - Marina Cugurra ATOS - Iván Martínez
Reviewer(s)	ETA - Marina Cugurra ATOS – Iván Martínez

Abstract	Deliverable D6.6 is the final deliverable of WP6 of the DataVaults project. It providing a condensed list of recommendation and critical success factors that should be considered by organisations interested to scale up DataVaults (or any other relevant personal data sharing platform). They cover the organisational, legal and ethical as well as the technical dimensions of such platforms.
Disclaimer	The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. © Copyright in this document remains vested with the DataVaults Partners

Executive Summary

Deliverable D6.6 is the final deliverable of WP6 of the DataVaults project. It aims to document and provide information for scaling up the DataVaults (or any other personal data sharing platform that acts as a brokerage infrastructure bringing together data owners and data seekers), by outlining a condensed list of recommendation and critical success factors that should be considered by interested organisations. As such, the deliverable tries to touch aspects that belong at organisational, legal and ethical as well as the technical level.

The main target audience of this deliverable is as follows: organisations that are interested to deploy their own instance of DataVaults, as well as organisation that are interested to join an existing DataVaults instance. These stakeholders will be the driving force dealing with the deployment of DataVaults instances and will be responsible for maintaining them. At the same time, Data Owners will be of course the audience to be targeted to scale up the DataVaults deployments in terms of utilisation, nevertheless we consider that actions shall start from the side of Data Seekers, to place the foundations for making a DataVaults instance a successful deployment.

For these organisations, regardless of if they want to act as data brokers or as simple users of the platform (as “Data Seekers”), certain pre-requisites shall be fulfilled. These concern their readiness at organisational, technical as well as legal level, as data management activities, and especially of personal data, come with a high number of responsibilities and compliance needs. These have to be a-priori solved, by pushing these organisations to adopt solutions that are in a position to guarantee the efficient, effective and legislation compliant operation of the whole infrastructure

Aside of those pre-requisites, organisations then have to work at two different levels to enable their deployment to scale up. The first considers methods to attract end users (e.g., Data owners), which are the ones that are the driving force of the platform, as it is their data that makes the platform rich and attracts in turn other Data Seekers and Data Owners.

The second level of actions concerns the factors that deal with how the legal and ethics aspects shall be tackled, as well as technical and organisational best practise guidelines, that have been recorded during the demonstration period of the project by the different demo partners. In terms of legal and ethics considerations, the main legal and ethical challenges that arisen in DataVaults are discussed, alongside with the technological methods that have been used to tackle them. For the organisational and technical dimension, recommendations are provided to potential adopters that have been distilled by the partner’s experience during the development, operation, utilisation and evaluation of the overall solution.

Table of Contents

List of Figures

Terms and Abbreviations

DID	Decentralised Identifier
Dx.y	Deliverable x.y
EC	European Commission
EDPO	Ethics and Data Protection Officer
eID	Electronic Identity
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
MVP	Minimum Viable product
SSI	self-sovereign identity
UX	User Experience
WPx	Work Package x

1 INTRODUCTION

Deliverable D6.6 is the final deliverable of WP6 of the DataVaults project and is composed by analysis the experience of the project during the demonstrator phase.

1.1 SCOPE OF THE DELIVERABLE

As the final deliverable of the demonstration work package of the project, the aim of the document at hand is to provide to interested stakeholders a view on how organisations that are willing to adopt DataVaults should work towards making the operation of the platform a success.

In any case, the content in this deliverable can be also seen as general guidelines for the adoption and operation of any other personal data sharing platform, which includes functionalities similar as DataVaults, as the main concept behind those would be the same. It is however noted that the present deliverable does not constitute a thorough analysis of success factors and of legislation and operation principles for any other personal data platform.

In this context, the scope of this deliverable is to:

- Identify the main areas which an organisation that would like to exploit DataVaults as a broker or as a Data Seeker should invest in, at the organisational, legal, technical as well as human resources level.
- Discuss certain lessons learned from the DataVaults experience that should be taken into consideration when building such platforms.
- Expose some of the benefits recorded at high level by the different stakeholder groups, based on the experience recorded in the demonstrators of the project.
- Summarise all the above into a list of critical success factors that should be carefully considered when aiming to operate the DataVaults solution (or any other similar personal data sharing platform).

1.2 DOCUMENT STRUCTURE

The deliverable at hand is structured as follows:

- Section 2 presents the main audience to which this deliverable is addressed.
- Section 3 discusses briefly, from a theoretical viewpoint what scale-up activities are and how they can be performed.
- Section 4 focuses on the pre-requisites that should be fulfilled by organisations that are interested to operate as DataVaults brokers or simply as Data Seekers
- Section 5 discusses means that can be used to attract audience, as identified by the demonstrator partners, based on their experience during the project
- Section 6 provides some lessons learned during the demonstration phased, presenting the knowledge gained from an ethical and legal methodology perspective, as well as the benefits for the different stakeholders as recorded in the demonstrators.
- Finally, Section 7 concludes the document.

2 DATAVAULTS TARGET AUDIENCE AND SERVICES

DataVaults comes as a service that is able to bridge the gap between data providers and data consumers of personal data, in a novel manner where ICT plays a key role in the automation, trustworthiness and efficiency of the whole process.

In essence, the platform is in a position to generate a new relationship of trust, which can evolve into a business relationship between individuals and personal data consumers, allowing the former to execute full control on their data and being the deciders on how, when, with how and under which terms they can be shared, and offering to the latter group the possibility to consume data from populations of interest, that are more accurate and are representative, as done with other existing, manually executed activities.

In this context, DataVaults comes as a solution for a Personal Data Marketplace and in principle concerns two main groups:

- Individuals (called **Data Owners**): These are the driving force of the overall concept as these are the subjects that utilise the platform to collect at a single point many of their data (much of which is generated online in third party services and is collected via connectors), and then set the rules on how their data can be shared and under which terms. The
- Data Consuming Organisations (called **Data Seekers**): These are organisations (or sole traders) which are interested in acquiring personal data to improve their services and build data-drive services and products. This group, with the help of DataVaults is able to directly acquire personal data from individuals, without going through third parties that are already collecting these data as part of their services and that are handing over this intelligence to other organisations for a fee.

As the platform is conceived and has been demonstrated during the project, registration to the platform is a very easy task and is subject of both groups agreeing with the different legal terms that are relevant to the management of personal data, ethics, and the use of such data, adhering to all GDPR guidelines. The rationale behind this approach is directly linked with the fact that the more data owners and data seekers are registered to the platform, the richer the "data lake" of DataVaults can be made, and the more data transactions can happen.

Of course, onboarding to the platform can be constrained as well, as there exist mechanisms that are able to enable moderated registration requests (without revealing the identity of users), however this feature is provided in case an instance of the platform needs to be deployed and run in "isolation" mode, allowing only a specific group of users to use it (for example if it is to be used for running campaigns and conducting analysis between followers/customers of a specific brand which will operate the platform, if it is to become a key data retrieval point regarding citizens data in case the platform is hosted by a public administration, etc.)

In any of those case, we consider as the main target audience for adopting the DataVaults solution (e.g., the exploitation/business target group) to be that of Data Seekers, and the rest of this chapter is focused on this group.

2.1 MAIN TARGET AUDIENCE

As identified in the previous section, the main target audience which is concerned for adopting DataVaults is that of Data Seekers, as it is expected that it is these organisations which will be able to invite their customers/followers/audience to the platform to engage with them in a win-win personal data trading activity, that respects legislation, ethics and privacy, as evangelised in the DataVaults concept.

Of course, it is not expected that all Data Seekers will carry onboard their audience, as many of such organisations might join once they see a critical mass of individuals already operating over the platform, however for reaching this state it is important to identify the competences and expertise needed for early adopters of the solution and provide some guidelines on how to better prepare for operating withing the DataVaults context.

In this direction, we distinguish early adopter Data Seekers in two major groups:

- Organisations that are interested to **deploy their own instance of DataVaults**. This can happen either in "isolation mode" (e.g., working only with their audience group) or by offering a publicly accessible instance (in terms of making it open for Data Owner registration only). Such organisations are in most of the cases ones that possess a large audience group which they would like to reach via the platform and enjoy more accurate and efficient results. In both operation modes, such organisation can be also become "data brokers" in case they allow the registration (moderated or not) of other organisations into the platform as Data Seekers. In such scenarios, these data brokers will be able to generate new revenue streams as they could define certain service fees and commissions for the different transactions or services offered to other organisations through the platform. In any of the above cases, the deployment, operation and maintenance of the platform by such an organisation is calling for the presence of a set of specific skills (not all of them of IT-nature, as the pure technological infrastructure can be also handled by a technical provider partner), and the following sections present an overview of the main competences/roles that such an organisation should possess in order to successfully operate such an instance, both from the technological as well as from the business world perspective.
- Organisation that are interested to **join an existing DataVaults instance**. This is the case of organisation that want to exploit the personal data lake that exist in one of the DataVaults deployments and that become followers of the approach and users of the system, presumably paying for the services they enjoy to an organisation that is hosting the platform, apart from paying some fees to Data Owners to acquire their data. It is self-standing that in this case, the entry barrier to become a Data Seeker in an already deployed DataVaults platform is dramatically lower that this of running and sustaining an instance (as in the previous bullet). Nevertheless, in order to fully exploit the potential of the services provided by DataVaults, it is essential also for such stakeholder to possess some skills, while it is very important for the whole ecosystem that these Data Seekers, motivated by the quality of service they will get, are pitching the idea to individuals to join the platform and to further enlarge the personal data lake offered.

2.2 SERVICES OFFERED BY THE DATAVAULTS PLATFORM

Before discussing the main pre-requisites for any organisation that would like to operate the DataVaults solution, or simply become a user in an existing instance, this section summarises the main services that are currently offered by the platform to both Data Seekers and Data Owners, for any interested party to better understand what can be done with the platform.

As such, the next figure provides a high-level list of features offered to both these groups. For more information on the specific workflows and details on the different features are provided in the deliverables D3.3, D4.3 and D6.6

Features for Data Owners
Protect Personal Information while Sharing Data
Provide Demographics Profile Data
Allow Profile to be searchable based on Demographics
Fetch Data from External Web Data Sources
Fetch Activity and Location Data from Mobile App
Upload Files
Selected Data to be Shared and Sharing Policies
Anonymise Data to be Shared
Select Data to be part of aggregated Personas
Encrypt Datasets for extra security
Define Access Policies on Data
Define Data Prices
Sign Transactions
Receive and Answer Questionnaires
Negotiate over Data Prices
See Privacy-Risk Index based on Shared Data
Acquire Compensations by redeeming points

Features for Data Seekers
Search for available Data Sets
Acquire Datasets for a certain price
Negotiate with Data Owners for Dataset prices
Compile and Retrieve Questionnaires
Search for available aggregated data as Personas
Place Compensations in the Marketplace
Run Analyses on acquired dataset[1]

Figure 1: Data-Relevant Services offered by the DataVaults Platform

As it is seen in the table above, the services offered to Data Seekers are fewer than those offered to Data Owners, and this is logical as the scope of the platform is largely to allow Data Seekers to find and acquire personal data and then use them for their own causes, always based on the licenses and agreements that accompany those data. This approach is also encouraging Data Seekers to join the platform as it is fairly easy to use, including a set of specific and non-complex features that can be performed by anyone with basic knowledge of digital technologies

3 Scale-Up Roadmap

A scale-up roadmap for the DataVaults platform outlines the steps and milestones necessary to grow and expand the platform's reach, user base, and impact across various industries and organizations. A specific scale-up process is provided in this chapter, involving five phases, each with its own objectives, challenges, and milestones. Careful planning and execution during each phase are a basic requirement in order to ensure that the platform grows effectively, constantly expanding its reach.

In detail, the scale-up journey for the DataVaults platform can be divided into five key phases as shown in the next figure "Launch, Expansion, Growth, Global Expansion, and Continuous Improvement and Innovation". Each phase requires distinct actions and strategies to navigate the challenges and capitalize on the opportunities that arise during the scale-up process.

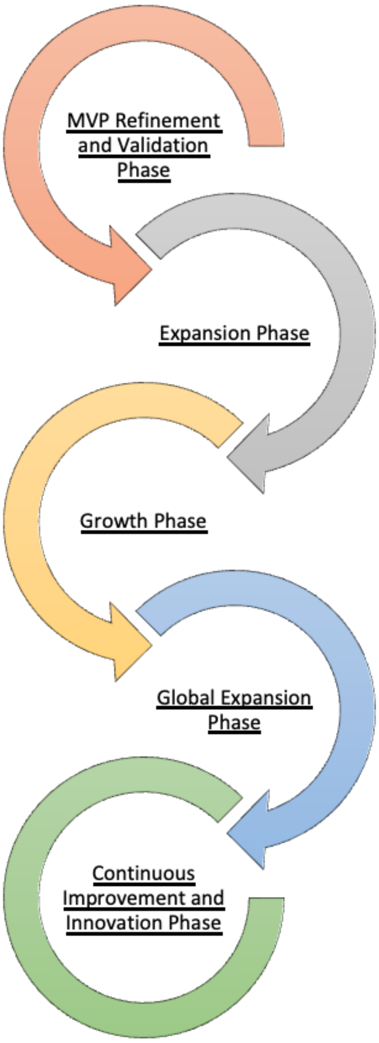
MVP Refinement and Validation Phase
Expansion Phase
Growth Phase
Global Expansion Phase
Continuous Improvement and Innovation Phase

Figure 2: DataVaults Scale-Up Journey

1. MVP Refinement and Validation Phase

   Given that the development and implementation of a Minimum Viable Product (MVP) that meets the data sharing and monetization needs of different target industries has already taken place in the framework of the implemented project activities, the primary goal in this phase is to validate its value proposition and assess its performance, user experience, and security features. Key objectives during this phase are to validate the platform's value proposition, identify potential improvements, and establish a solid foundation for future growth. This phase includes:

   - Piloting the platform with early adopters (including but not limited to the pilot partners)
   - Gathering feedback and iterating on the MVP to improve its functionality, user experience, and data security.

2. Expansion Phase

   The Expansion Phase remains focused on extending the platform's reach to new industries and user segments. This includes the following:

   - Conduct further market research to identify potential industries that can benefit from the platform's features and adapt the platform to cater to the specific needs of these industries.
   - Continue to develop and strengthen partnerships with key industry stakeholders, influencers, and thought leaders to facilitate the introduction of the platform to new markets.

- Invest in marketing and promotional efforts tailored to each industry to increase platform awareness and adoption.

3. Growth Phase

   In the Growth Phase, the emphasis is on accelerating user adoption and expanding the platform's feature set, including the following:

   - Refine and optimize the onboarding process, enhance customer support, and proactively address user concerns to ensure a seamless experience.
   - Continuously monitor user feedback and industry trends to identify opportunities for platform improvements and feature additions.
   - Scale up marketing and promotional activities to drive awareness and adoption among potential users and partners.
   - Collaborate with complementary platforms, services, or products to boost the platform's growth.

4. Global Expansion Phase

   The Global Expansion Phase continues to involve entering new geographic markets with high demand for privacy-preserving data sharing and monetization solutions. These mostly refers to the following:

   - Conduct market research to understand the legal, cultural, and technological landscape of each target region.
   - Adapt the platform to meet the specific requirements of these new markets, build relationships with local partners, and design localized marketing campaigns and promotional activities.

5. Continuous Improvement and Innovation Phase

   The Continuous Improvement and Innovation Phase still focuses on ensuring the platform remains competitive, relevant, and effective in meeting user needs. The related activities include:

   - Regularly monitor platform performance, user feedback, and industry trends to identify areas of improvement and potential innovation.
   - Stay up to date with technological advancements, regulatory changes, and market shifts.
   - Introduce new features and updates regularly to maintain a strong value proposition and keep users engaged.

While considering the platform version(s) developed during the project as the MVP and adjusting the scale-up phases accordingly, the DataVaults platform can effectively grow its user base, expand into new industries and markets, and establish itself as a leading solution for privacy-preserving data sharing and monetization. In the next sub-sections, we discuss the pre-requisites and the critical success factors that can lead to the successful scale-up of the DataVaults solution, to facilitate potential adopters of the solution (which is provided as open-source prototyped software), to continue the work and drive the creation of a production-grade ecosystem for personal data sharing.

4 PRE-REQUISITES FOR ORGANISATIONS WORKING WITH DATAVAULTS

In this section, we discuss the different pre-requisites necessary for organisations that want to be engaged with the DataVaults solution, indicating certain aspects in terms of technical, data and legal requirements, as well as organisational ones.

In the sub-sections that follow, the distinction between the case where an Organisation is using DataVaults as the provider of the services (and potentially also becoming a data broker), and that of the organisations acting as simple users of the platform (as a Data Seeker).

4.1 ORGANISATIONAL PERSPECTIVE

4.1.1 Organisations that want to Deploy DataVaults

In case an organization is willing to become the operator of a DataVaults instance there are a number of requirements that need to be satisfied in order to allow the efficient and effective operation of the service, as in principle the organization will become a digital service provider.

As such the organization should fulfil the following requirements:

- Organizational Structure and Management: Establish a clear organizational structure and management hierarchy for the DataVaults platform, outlining the roles and responsibilities of each team member. This structure must provide a clear framework for decision-making, coordination, and communication within the organization, facilitating efficient and effective operation. It’s proposed a dedicated IT team to be formed, able to maintain and debug the service, working on the deployment of the components and the maintenance by making the necessary updates to the libraries, code and infrastructure.
- User support: A user help desk shall be organised in order to support user requests and offer a help line to solve problems that users might face,
- Marketing and Promotion Strategy: Develop a comprehensive marketing and promotion strategy to raise awareness about the DataVaults platform and attract users, partners, and stakeholders. This strategy should include a mix of online and offline marketing activities, such as content marketing, social media advertising, public relations, and participation in industry events or conferences. It’s proposed to tailor the marketing approach to the specific needs and preferences of each target audience and industry segments to maximize its impact.
- Formation of a Software Development Unit: Establish a dedicated software development unit within your organization to focus on developing new data connectors and increasing the data ingestion capabilities of the DataVaults platform. This unit should comprise skilled developers, engineers, and data experts who can work together to identify, design, and implement data connectors for various data ingestion scenarios. The development unit should also collaborate closely with other teams within the organization, such as marketing and customer support, to ensure the platform evolves to meet user needs and market demands effectively.
- Data Management, Privacy and Data Protection Compliance: Development of a comprehensive data management plan for the data that will be used over the

platform. The plan as well as all the policies of the organisation shall comply with relevant data privacy and protection regulations, starting from the General Data Protection Regulation (GDPR), setting strict rules for the collection, processing, and storage of personal data and require organisations to implement appropriate security measures to protect users' data. It's crucial to ensure that the DataVaults platform adheres to these regulations and any other applicable laws in the jurisdictions in which it operates.

- Security Infrastructure and Policies: Develop and implement robust security infrastructure and policies to protect users' data and maintain the platform's integrity. This may include encryption protocols, secure data storage solutions, access control mechanisms, and regular security audits. It's important to establish clear security policies and guidelines that govern the handling and processing of data, as well as a comprehensive incident response plan in case of a security breach.
- Intellectual Property Protection: Secure the necessary intellectual property rights for the DataVaults platform. This will help in safeguarding the platform's unique features, functionalities, and brand identity, ensuring that competitors cannot exploit them without permission.

Organisations that wish to operate as data brokers, in addition to operating the DataVaults platform, must meet several additional requirements to ensure compliance with industry regulations, protect user data, and maintain transparency, like:

- Formation of a Customer Relations Unit: Establish a dedicated customer relations unit within the organisation to manage the onboarding and ongoing interaction with data seekers who will become customers of the data broker. This unit should comprise skilled professionals who can effectively communicate the value proposition of the data brokerage services, address customer inquiries and concerns and provide personalized support to facilitate a smooth onboarding process and maintain strong relationships with data-seeking organisations.
- Design of a Financial Service Strategy and Pricing Model: Develop a comprehensive financial service strategy and pricing model for the data brokerage services, which covers fees for onboarding organisations and commission rates for transactions facilitated through the platform. This strategy should take into consideration factors such as market trends, the value of the data being shared, the costs associated with operating the DataVaults platform, and the organisation's competitive positioning. As far as it concerns the pricing model, it should be transparent, fair, and flexible enough to accommodate a range of data seekers and transaction types, while also ensuring that the organization generates sufficient revenue to support its operations and growth.
- Transparency and Disclosure: As a data broker, the organisation must maintain transparency about its data collection, processing, and sharing practices. This includes disclosing the types of data collected, the sources of the data, and the purposes for which the data will be used. The organisation may need to update its privacy policy and terms of service to reflect these practices and ensure they are easily accessible to users.

- Opt-Out Mechanisms: The organisation must provide users with clear and easily accessible mechanisms to opt-out of having their data collected, processed, or shared. This may include a dedicated interface as well as clear instructions on how to submit opt-out requests. Comply with users' opt-out requests promptly and ensuring that their data is not shared with third parties without their consent is an extremely crucial requirement.

4.1.2 Organisations joining a DataVaults Instance as Data Seekers

When it comes to organisations that act as data seekers, no specific requirements are imposed from this perspective. In this case, such organisations are simple users that seek and acquire datasets, in a more robust and trusted manner as they are already doing, and in principle no change is required in their organisational structure or procedures for these activities, as it can be seen as a data procurement action that is realised over a different channel (e.g., via the DataVaults platform

4.2 Technical Requirements Perspective

This section discusses the technical requirements necessary to run the DataVaults service. These refer to technical infrastructure (hardware) that is necessary to run the different components of DataVaults.

The software components of the DataVaults platform are accessible under open-source licenses from the project's open source Gitlab repository using the following link: https://gitlab.com/datavaults

4.2.1 Organisations that want to Deploy DataVaults

The technical infrastructure that is necessary for the Deployment of the DataVaults solution is the following:

- 1 Server for hosting the DataVaults Cloud Platform with the following minimum requirements:
  - 4 vCPUs
  - 32 GB of RAM
  - At least 128GB or usable Disk Space
  - Operating System: CentOS

- 1 Server for hosting the DataVaults Personal App with the following minimum requirements:
  - 4 vCPUs
  - 32 GB of RAM
  - At least 128GB of usable Disk Space
  - Operating System: CentOS

- At least 2 Servers for running the DataVaults Blockchain Network with the following requirements[1]:

[1] In order to ensure high availability on the blockchain, at least 8 VMs with the above specifications are necessary, as described in D5.6.

  - 2 vCPUs
  - 16 GB of RAM
  - At least 648GB of usable Disk Space
  - Operating System: CentOS

These requirements are the minimum requirements required to run the service with a population of 1.000 Data Owners and 50 Data Seekers, and as the number of Data Owners and Data Seekers are growing, these servers should be scaled (and new blockchain nodes to be added).

In case the organisation is interested to utilise the SEAS analytics system, then the only requirement is to set up and run SEAS via docket. SEAS is completely dockerised and includes a docker compose file that facilitates deployment. The user may quickly update the SEAS settings based on Jupyter [1], MLflow [2], and Apache Superset [3] for local configuration by modifying the given local environment files associated (.env).

The hardware requirements for SEAS are as follows:

- 4 VCPUs (8 cores recommended)
- 16GB of RAM (maximum 32GB
- At least 20GB of usable Disk Space (40-50GB recommended)

4.2.2 Organisations joining a DataVaults Instance as Data Seekers

In case of an organisation acting as a simple user (as Data Seeker), no specific technical infrastructure is necessary.

In case the organisation is interested to utilise the SEAS analytics system, then the requirements for its deployment are the same as in the previous sub-section.

4.3 Data Requirements Perspective

From the Data requirement perspective, there is no specific requirements for organisations that want to operate over DataVaults, as Data Owners can utilise the existing available connectors provided by the project, or directly upload files.

4.3.1 Organisations that want to Deploy DataVaults

In order to improve the data lake available to Data Seekers, DataVaults operators might consider developing new connectors to allow Data Owners to fetch their data and make them available through the platform. To do so, there needs to be a careful investigation on the data sources that will be connected to DataVaults, focusing on the following aspects:

- Verification that the data can be legitimate extracted from the third-party system, by analysing the terms of these services
- Verification that the data to be extracted can be used for the causes of DataVaults
- Verification that the extraction process is based on user (Data Owner) credentials, as this action should be invoked by them, and the DataVaults operator should not be able to fetch those data

- Validation of the API response provided by the third-party service is in accordance with what is expected from the DataVaults side and that performance is acceptable.

Following these checks, an operator can move on with the development and the integration of a connector into the system, allowing the Data Owner to fetch himself the data and store it into the system.

4.4 Legal and Ethical Requirements Perspective

As far as it regards the legal and ethical requirements for organisations onboarding DataVaults, first of all it is key to identify and analyse the legal framework specifically relevant for the application context, covering the EU level instruments and the national, local and sector-specific instruments.

Regarding the European sources, the following are or might be relevant:

- Privacy and Data Protection Legislation, paying special attention to i) GDPR ("General Regulation on data protection" 2016/679); ii) to the "ePrivacy Directive" (Directive 2002/58/EC on privacy and electronic communications), which replaced the Directive 97/66/EC and was partially amended by Directive 2009/136/EC; and iii) to the national legislations in the countries where DataVaults is going to be applied;
- the Human Rights Law, including the European Convention of Human Rights and the Charter of Fundamental Rights of the European Union;
- the Ethics and Soft Law (quasi-legal instruments), including the European Courts' case law. As explained in D2.2, these instruments, though not necessarily legally binding, usually are very helpful, especially in filling the gaps of the legislation, in identifying safeguards, boundaries and obligations to ensure the legitimacy and fairness of the new technologies and in identifying the balance between competing interests on a case-by-case basis. An example is the European Data Protection Supervisor's Opinion 7/2015 "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability";
- the vast reforms and regulatory instruments under development, such as the Data Governance Act, the Digital Service Act, the Digital Market Act and the Proposal of Regulation on Privacy and Electronic Communications (E-Privacy Regulation).

It is recommended that in any case of future uptake of DataVaults, those involved at that time should check whether these pieces of legislation are relevant for DataVaults adoption in their specific context and therefore whether if some additional legal/ethical requirements arise, besides those identified in D2.3 and applicable in the post-project phase:

- Fairness and Lawfulness
- Purpose limitation and legitimate aim
- Data minimization
- Data accuracy
- Integrity and confidentiality
- Storage Limitation

- Transparency
- Privacy and Data Protection by Design and Privacy by Default
- Avoidance of discrimination (including social sorting) and of harm
- Informed Consent
- User Control
- Data subject’s rights
- Enforcement
- Privacy Notice
- Data breaches
- Accountability
- Record of processing activities
- Data Protection Impact Assessment
- Application scrutiny to local/national boards if required by national legislation concerned
- International Data Transfer
- Technical and organizational measures
- User and data protection friendly User Interface
- Measures in case of profiling
- Appointment of Data Protection Officer
- eIDAS Obligations

More information regarding these requirements can be found in D2.3. Additional recommendations, take-aways and hints relevant for the future adoption of DataVaults are provided in Sect. 6 of this document.

4.5 IT Skills/Human Resources Requirements Perspective

To operate or work with DataVaults, a specific set of IT skills are required. Again, in this sub-section we distinguish Data Seekers from DataVaults Operators from, as the latter group clearly requires having a more elaborate set of skills to run the platform.

4.5.1 Skills required for Data Seekers

Data Seekers who wish to use the DataVaults platform should possess a solid foundation of skills, as they may need to perform various advanced IT operations and navigate through different aspects of the platform. Although basic user journeys related to data querying, selection, and acquisition are generally simple, some platform features require users to understand fundamental concepts and terms related to data privacy, security, and analytics.

In this context, Data Seekers should have a good grasp of the following aspects:

- Data Privacy and Security Principles: Users should be familiar with key data privacy and security concepts, such as anonymization, pseudonymization, and digital personas, which help protect sensitive information when sharing data with other parties.

- Data Encryption Techniques: Data Seekers should understand the basics of data encryption, including asymmetric encryption and attribute-based encryption, to ensure the secure transmission and storage of data.
- Data Licensing Knowledge: Users should be able to comprehend and interpret data license information to ensure compliance with the terms and conditions set by Data Owners when sharing data.
- Blockchain Wallet Operations: Data Seekers should be familiar with the basic principles of blockchain wallets, including how to set up, manage, and secure wallets for use with the DataVaults platform.
- Service Deployment Fundamentals: Users should have a basic understanding of service deployment, such as setting up service endpoints and deploying virtual machines (VMs), to ensure smooth integration with the platform and access to its features.
- Data Analytics Skills: Data Seekers should be proficient in using data analytics tools, such as Jupyter notebooks and visualization engines, to process and interpret the data acquired through the platform effectively.

By equipping themselves with this knowledge, Data Seekers can maximize their ability to use the DataVaults platform effectively, ensuring they can access, acquire, and analyse the data shared by Data Owners while maintaining the privacy and security standards required.

4.5.2 Skills required for DataVaults Operators

For organisations operating the DataVaults platform, the required skills can be grouped into three distinct categories:

1. Data Seekers within the Organisation: These individuals will work as users of the platform from the organisation's side. The necessary skills for this group are essentially the same as those outlined for Data Seekers in the previous section, ensuring they can effectively navigate the platform, acquire data, and perform analytics.
2. Platform Management Personnel: This group of individuals will be responsible for supporting users, both Data Owners and Data Seekers, in case the organisation operates the platform as a broker. They must have an excellent understanding of the platform from both the Data Owner and Data Seeker perspectives, enabling them to provide comprehensive support. In addition, they should possess strong communication and problem-solving skills to address user inquiries and issues effectively.
3. Deployment, Maintenance, and Improvement Team: This group comprises technical professionals responsible for the deployment, maintenance, and enhancement of the platform. They form a software development team that needs to work on the platform's code and ensure its smooth operation. The skills required for this group are as follows:
   - Expertise in Software Development, including:
     - Backend Technologies (depending on components): Java, Python, Node.js, Go, RabbitMQ, Linux
     - Frontend Technologies (depending on components): Vue.js, Angular, Tailwind
     - Deployment Technologies: Docker, Ansible

  - Blockchain Technologies: GoQuorum
- Proficiency in Software Integration: Individuals should have experience integrating various software components and systems to create a seamless user experience.
- Comprehensive understanding of Software Licensing Terms: Team members should be well-versed in software licensing terms to ensure compliance with relevant regulations and requirements.

By ensuring that individuals within each group possess the necessary skills, organisations can effectively operate the DataVaults platform, support users, and maintain and enhance the platform's functionality to meet evolving user needs and market demands.

4.5.3 Skills required for Data Owners

Although in the previous sub-sections Data Owners were not discussed, as they are not the target audience regarding the adoption of the platform from a business interest perspective (despite them being the main force to make the platform successful), in this section the skills that are required by this group are discussed.

In principle, the skills required to work with DataVaults are adequate if they are of basic level, as the UX of the Personal App is highly intuitive and easy to use, even for not deeply IT-literate persons. However, to fully exploit the services of the platform, Data Owners need to be literate around aspects relevant to data privacy and security (such as anonymisation, encryption, TPM signatures, etc). Absence of such knowledge is not blocking for fetching and sharing data, and to allow Data Owners to select advanced features without requiring them to invest a lot of time to learn about them, the platform includes easy to understand templates that provide pre-selected options for different privacy and security data management levels.

5 HOW TO ATTRACT INDIVIDUALS TO DATAVAULTS

DataVaults is a ground-breaking platform that empowers individuals to take control of their personal data, share it securely, and monetise it based on their preferences. To maximise the platform's potential, it is essential to attract a diverse range of individuals to use DataVaults.

5.1 ENGAGEMENT STRATEGIES

Next, different strategies for engaging various groups of individuals are outlined, starting with smaller groups, and scaling up to larger populations, while providing incentives, both monetary and non-monetary, to encourage participation.

i. Engaging Small Groups

Initially, the focus should be on engaging small groups of individuals who may have a heightened interest in the platform due to their data privacy concerns or potential for data monetization. These groups may include niche communities, such as tech enthusiasts and freelance professionals who work with data.

Strategies for engaging these groups may include:

- Hosting targeted events, such as webinars and workshops, to introduce the DataVaults platform and demonstrate its benefits. These events can be tailored to specific audiences, such as industry professionals, data privacy advocates, or the general public.
- Engaging with online communities and forums where these individuals congregate, sharing informative content, and addressing questions and concerns.
- Developing partnerships with influencers and thought leaders within these communities to promote DataVaults and provide testimonials.
- Utilising various marketing channels, such as social media, email newsletters, and online advertising, to reach potential users and educate them about the platform's benefits.

ii. Scaling Up to Larger Populations

As the platform gains traction within smaller groups, efforts should be made to scale up outreach and attract larger populations.

Strategies for achieving this may include:

- Implementing a referral program that rewards existing users for inviting new individuals to the platform.
- Collaborating with larger organizations, such as public administrations, universities, companies, and NGOs, to introduce DataVaults to their networks and encourage adoption.
- Investing in targeted marketing campaigns, leveraging social media platforms, search engine advertising, and content marketing to reach a broader audience.

- Organise corporate events: Host events to demonstrate the platform's functionalities, address user concerns, and stimulate interest in personal data management. Additionally, invite guest speakers and experts in the field of data privacy and security to provide additional insights and perspectives, further enriching the event experience for attendees.
- Organise hands-on training and interactive demonstration workshops to help users understand the platform's features and benefits. These workshops can be tailored to specific audiences, such as industry professionals, data privacy advocates, or the general public.

iii. Providing Incentives

Offering incentives, both monetary and non-monetary, can significantly boost user engagement and attract more individuals to DataVaults.

Indicative incentives may include:

- Monetary incentives, such as sign-up bonuses, referral rewards and exclusive promotions for users who share their data.
- Non-monetary incentives, such as access to premium features, exclusive content, or discounts on partner products and services for users who actively participate in the platform.
- Gamification elements, such as badges, leader boards, and challenges to encourage users to engage with the platform and compete with their peers.

iv. Fostering Community and Ongoing Engagement

To maintain user interest and ensure long-term participation, it is crucial to create a sense of community and provide ongoing opportunities for engagement.

Strategies for fostering community may include:

- Offering a user-friendly platform interface that encourages interaction, feedback, and collaboration among users.
- Developing a content strategy that provides valuable, informative, and entertaining content to users on a regular basis, such as blog posts, newsletters, and social media updates.
- Hosting periodic events, such as webinars, meetups, and conferences, to provide networking opportunities, showcase new platform features, and celebrate user achievements.

5.2 EXPERIENCES FROM THE DEMONSTRATORS

In the present section, we outline selected experiences from demonstrators that have successfully attracted users to the DataVaults platform. We discuss what has worked well, what has not, and how these insights can inform future strategies for engaging individuals with DataVaults.

i. **Challenges in Initial User Engagement - Insights from the Municipality of Piraeus:**

The Municipality of Piraeus found that using social channels, initial meetings, and visits to local shops and the cruise terminal did not result in the anticipated number of user enrolments. This was mainly attributed to the platform's maturity as a "research" project, as this is attributed to people's reluctance to spend time on a local pilot of a platform that was not yet widely accessible, especially during their vacations (when referring to tourists). Future strategies should focus on timing the outreach efforts to coincide with the platform's higher maturity and targeting users during periods when they are more likely to engage.

ii. **Maintaining User Interest - Insights from MIWenergia:**

MIWenergia faced challenges in keeping participants involved and interested in the project. The initial demonstration activities were hampered by the platform being at a beta stage which came with some bugs that in some cases prevented users from sharing their data and fully testing all the platform's features. This caused them to lose interest. Additionally, the absence of rewards at the beginning of the project meant there was no extra incentive for users to engage with the platform. To address these issues, future implementations should ensure that technical issues are resolved promptly, and incentives are in place from the beginning to maintain user interest and engagement. As it was found out, it was easier to attract and retain users closer to the end of the project, where the platform was clearly more mature, with an improved UX, and rewards were made available over the platform.

iii. **Importance of Data Source Availability - Insights from Prato:**

Prato's pilot project found that the availability of data sources was crucial for user adoption. Although all basic functionalities were sufficiently tested, the lack of a wider variety of data sources represented a limitation. The unavailability of certain social channels, such as Facebook, due to licensing limitations hindered the project's effectiveness as users could only access data from Twitter, which is not widely adopted by the local population. In future deployments of the DataVaults platform on a large business scale, the availability of diverse data sources should be addressed in a structured way, as this will represent a real added value for the adoption of the DataVaults solution by a large number of users.

5.3 Conclusions On Attracting Individuals to DataVaults

Throughout this chapter, we have explored various strategies for attracting individuals to the DataVaults platform, taking into consideration the experiences and lessons learned from demonstrators operated by the Municipality of Piraeus, MIWenergia, and the Municipality of Prato. By synthesizing these insights, we can draw several key conclusions to guide the future growth and expansion of DataVaults:

- **Targeted outreach and personal engagement are crucial for attracting users to the platform:** The pilot projects demonstrated the effectiveness of personal contacts, demo sessions, and training events in helping potential users understand the platform's benefits and value proposition. To scale up the platform, more extensive outreach through social networks, newsletters, and events will be necessary, while maintaining a personal touch and providing ample opportunities for user education and support.

- **Incentives play a significant role in user engagement and retention:** Offering rewards or incentives, such as discounts on energy bills, encourages users to participate and stay engaged with the platform. Future implementations of DataVaults should incorporate a variety of incentives, both monetary and non-monetary, to motivate users to share data and remain active on the platform.

- **Ensuring a seamless user experience is vital for maintaining user interest:** Bugs and technical issues during the early stages of the platform can negatively impact user engagement. Thorough testing and refinement of the platform before engaging users is essential to provide a smooth experience and prevent user attrition.

- **The availability of diverse data sources is key to attracting and retaining users:** Addressing licensing limitations and providing a wide variety of data sources in future implementations will help create a more robust and attractive platform that appeals to a broader range of users.

- **Organizational and legal prerequisites for organizations operating DataVaults should be clearly defined and communicated:** This has to be done with a focus on required skills, compliance, and organizational structures. This will help ensure smooth operations and encourage more organizations to adopt and operate the platform.

Apparently, the future success of the DataVaults platform hinges on effective user engagement strategies that address the lessons learned from the demonstrators, incorporating targeted outreach, incentives, a seamless user experience, diverse data sources, and clear organizational and legal prerequisites. By embracing these key principles, DataVaults can foster a thriving community of users that drives the platform's growth and long-term sustainability.

6 MAIN LESSONS LEARNT AND KEY TAKEAWAYS

During the operation of the DataVaults demonstrators, valuable feedback has been collected by the consortium that had to do both with the technical implementation aspects in order to follow an agile development processes to constantly improve the solution and add new, requested features, but also from the legal and business (scope) perspective which were very valuable to constantly check the overall proves against the legal and ethical requirements compliance, as well as to understand various off-platform activities that should be performed by the organisations that are backing-up the platform to increase the users base, the satisfaction of the audience and streamline the overall business operation of the platform.

The feedback collected has been used at the end of the project to extract some major lessons learned, which are presented below.

6.1 LEGAL COMPLIANCE AND ETHICAL SOUNDNESS OF DATAVAULTS TECHNOLOGY

This chapter provides an overview of the main lessons learnt and takeaway regarding the legal and ethical challenges that the design, deployment and operation of a personal data platform might imply and regarding how to effectively deal with them. They were elicited thanks to an in-depth literary review and analysis of the regulatory and ethical framework, including the reforms under development, combined with insights captured in relation to the design and development of DataVaults technology and hints generated during the experiences of conducting the DataVaults demonstrators.

Such lessons learnt and key takeaways were described in a comprehensive manner in the book chapter “Does Everything Conform to Legal, Ethical, and Data Protection Principles?” [4] within “Personal Data-Smart Cities: How cities can Utilise their Citizens’ Personal Data to Help them Become Climate Neutral”. The following paragraphs provides excerpts and the most valuable findings taken from such chapter, specifically focusing on the DataVaults Project.

6.1.1 Key Legal and Ethical Challenges and Technology-Enabled Opportunities to tackle them

This chapter provides an overview of the main lessons learnt and takeaways regarding the legal and ethical challenges that the design, deployment and operation of a personal data platform might imply and regarding how to effectively deal with them. They were elicited thanks to an in-depth literary review and analysis of the regulatory and ethical framework, including the reforms under development, combined with insights captured in relation to the design and development of DataVaults technology and hints generated during the experiences of conducting the DataVaults demonstrators. Such lessons learnt and key takeaways were described in a comprehensive manner in the book chapter “Does Everything Conform to Legal, Ethical, and Data Protection Principles?” within “Personal Data-Smart Cities: How cities can Utilise their Citizens’ Personal Data to Help them Become Climate Neutral” [5] book. The following paragraphs provides excerpts and the most valuable findings taken from such chapter, specifically focusing on the DataVaults Project.

The operation of a personal data platform might imply several legal and ethical challenges, for instance related to personal data management in terms of data collection, data sharing and processing, as well as to the potential trade-off between the need to maximize data utility whilst protecting human rights and preserving meaningful human control [6].

In this chapter some important challenges and trends will be highlighted regarding the tools and technologies aimed at facilitating secure and trustworthy data sharing, taking inspiration and extracts from the work and regulatory surveys conducted within the DataVaults project, in conjunction with insights from recent debates and the literature.

In view of strengthening the development and growth of the data economy also in relation to personal data it is key to foster the adoption of trusted and secure personal data platforms capable of handling back control over the use of personal data to individuals giving them actual benefits, not-necessarily financial.

In DataVaults the efforts were directed towards building a win-win data sharing ecosystem in order to unlock the social value of personal data, going beyond user consent for fostering individual human empowerment and flourishing, as well as the common good of society and businesses' interests. In alignment with the EC's vision of personal data sharing that includes benefits for all the actors in the value chain, trusted, secure and value generating data management and sharing platform for personal data should be encouraged to the extent that they allow stakeholders' collaboration for supporting their own goals and operations, as well as allowing further stakeholders, such as the local communities and local authorities, to offer new socially and environmentally sustainable solutions and business models. In this environment, on the other hand, the technologies should move to regain the trust of individuals when it comes to data sharing, letting the control in their hands for deciding how, how much and in which manner they would like to share their data, whilst at the same time guaranteeing their privacy and with adequate security levels, as well as ensuring fair share of the value that their data generates, also in case of secondary operations.

Following this vision, human-centricity was at the centre of the DataVaults technological developments and should also characterize their future operation when it comes to data sharing. Prioritizing human well-being and fundamental rights and putting people first in the data-driven economy are expected to contribute to rebuild public trust and, therefore, societal acceptance of such innovations. This is also aligned with the Communication "2030 Digital Compass: The European way for the Digital Decade". Its Vision for 2030 relies on empowered citizens and businesses: "the European way to a digitalised economy and society is about solidarity, prosperity, and sustainability, anchored in empowerment of its citizens and businesses, ensuring the security and resilience of its digital ecosystem and supply chains" with four cardinal points for mapping the EU' trajectory.

DataVaults cloud-based platform for personal data sharing moves forward towards fully embracing this vision and promoting the EU's fundamental values (including protection of privacy) and is therefore expected to contribute in the future to the creation of a single market for data that will ensure Europe's global competitiveness and data sovereignty. This is also expected to increase the amount of data made available for use in the economy and society, at the same time safeguarding individuals by effectively empowering them to exercise their rights with regard to the use of the data they generate and to decide at a granular level about what is done with their data, moving towards "personal data spaces".

6.1.1.1 The need to avoid Consent Fatigue and to develop a use a user-and-data-protection-friendly User Interface

According to the GDPR, consent must be given for the processing of personal data for one or more specific purposes. In case of new purposes, it is necessary to either get fresh consent specifically covering such new purpose or find a different legal basis for the new purpose.

Even when expressed through electronic means, the consent of a data subject should be preventive and unambiguous. It requires a statement or clear affirmative action of the data subject. For instance, these actions can consist of ticking a box in an online environment, the choice of technical settings for information society services, and any other statement or conduct clearly indicating the data subject's acceptance of the data processing activities.

In a personal data sharing platform, it is also necessary to ensure that, where consent is obtained through use of a service-specific user interface (for example, within the a given personal data app or the interface of an IoT device), the individual must be able to withdraw consent through the same electronic interface with undue effort and without detriment.

The EDPS Opinion 7/2015 outlines important challenges relevant to data platform entailing the sharing of personal data, which were considered during DataVaults' development. Such Opinion clarifies that in many big data environments "*individuals cannot efficiently exercise control over their data and provide meaningful consent in cases where such consent is required. This is all the more so as the precise future purposes of any secondary use of the data may not be known when data is obtained: in this situation, controllers may be unable or reluctant to tell individuals what is likely to happen to their data and to obtain their consent when required*".

The data collection and processing in such data platforms might be intended for multiple purposes and it is necessary to ensure the consent for all these purposes (Recital 32 GDPR).

Recital 43 GDPR casts doubt on an approach based on one single consent form, broadly formulated as pre-emptively covering different future business models of the data controller.

Globalized, generic consent for multiple vague purposes risk to be assumed as not freely given and the question that arises is whether separate consent and the need for several, broken down consent requests are appropriate.

This aspect was explored in the context of DataVaults, whilst also reflecting on the need to avoid 'consent-fatigue' of a data subject.

As acknowledged by the Article 29 working Party, a layered approach could be a possible solution, still providing all necessary information step by step and providing balancing means of user control, whilst being substantially different by the mere use of pre-ticked boxes: it is not necessary that the first layer of information is completely in-depth about the details of the processing. It is important to explore if, for most of the cases (though not applicable to the special categories of personal data of Art. 9 GDPR), an implicit consent (such as a shade going away after a few seconds and assumes "yes") could work, after the first general consent during the installation of the service. It should be likewise investigating which information needs to be given to the data subject in which layer.

Useful indications can be retrieved in the following GDPR Recitals:

1. Recitals 32, which clarifies that it can be a written statement, including by electronic means, or an oral statement, if the data subject's behaviour clearly indicates his/her

acceptance of the data processing. It is recommended that if the data subject's consent is to be given following a request by electronic means, such a request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided.

2. Recital 33, which states that, being often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection, data subjects should be allowed to give their consent to certain areas of scientific research (or parts of research projects) when in keeping with recognised ethical standards for scientific research. “Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose”.
3. Recital 42, which states that *“…For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended. Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment*”.

DataVaults consent management policies are functional to ensure that the consent is:

1. “granular”, capable of providing distinct consent options for distinct processing operations; For each sharing operation, the technical platform is providing a new set of available option, thus allowing distinct consent provisioning.
2. specific to “one or more specific” purposes, ensuring that the data subject has a choice in relation to each of them; In the DataVaults system the data subject can decided for the different data he wants to share on the licensing terms that will accompany them and dictate their usage.
3. freely given, in the sense that the data subject should be able to exercise a real choice, without risk of deception, coercion, intimidation, or significant negative consequences if he/she does not consent.
4. informed, being the provision of information to data subjects prior to obtaining their consent necessary to enable them to understand what they are agreeing to, make informed decisions, and exercise control and, in general, their rights (including to withdraw their consent). As noted, a layered approach could help in this regard; For this purpose, the DataVaults platform provides the complete set of chosen sharing options that the data subject has selected, for the latter to verify his actions, while at the same stage (before the actual sharing of the data) the data subject is inform about his risk exposure metric.
5. separate from other terms and conditions.

In strict correlation with the informed consent topic, another issue arises: the personal data platform or application also requires adopting user and data protection friendly User Interface (UI), capable of facilitating as much as possible the user control features and consent management in an easy manner. It should be capable of collecting consent and constraints/restrictions, providing appropriate options for user information and control, thereby enabling the data subject to easily consent and exercise his/her rights set forth under data protection legislation, at national and European level. Consent is provided by the data subject while joining the platform, while all sharing actions are accompanied with a preview of the selected options by the user, while some templates have been implemented to easily allow the user to consent to data publishing by enforcing either maximum, or minimum data protections schemes.

An important element to consider is the wide range of data sources and to pay special attention in case where it includes sensitive information in the sense of Art. 9 GDPR.

In this case the consent has to be "explicit", in case of processing of special categories of data (this also applies to profiling activities, for instance). Though in many cases the term "explicit" could be interpreted as given in writing with a hand-written signature, in digital or online context like DataVaults, a data subject may be able to issue the required statement with other modalities (such as by filling in an electronic form, or by using an electronic signature). This is done by the data subject being the one who selects what data he wishes to share, and by him setting the conditions under which the data will be shared, and with whom. A filter on those data categories could allow the UI to distinguish between consent requests on “normal” personal data and those involving sensitive data. In the future, as an extension of the DataVaults functionalities after the end of the project, it could be investigating whether introducing functionalities for automatically detecting when sensitive data (or particular subset of sensitive data, for instance in the health care demonstrator) is collected, using machine learning techniques or other techniques and filtering such data.

The following challenges were considered and addressed in DataVaults:

- managing consent in a fine-grained way (including, for instance, partial granting or withdrawal of consent in some circumstances). This is supported by the platform at the registration process, while the deletion process when invoked is automatically also removing all data that is belonging to the data subject, apart from the data he has already agree to share to another party.
- managing the own data and exercise data subject’s rights in an easy way, for instance as regard adding, deleting, and rectifying personal data, and also including the possibility to access additional information in case of a data breach. This is supported by the DataVaults Personal App where the data subject can add, edit or remove the data he has uploaded to the platform.
- ensuring data portability and exporting the own personal information. This feature is available by the DataVaults platform, as all data can be exported either in CSV or in their original format.

6.1.1.2 The risk-based approach and the risk-exposure dashboard

Within a data sharing ecosystem, it is advisable, in relation to ethics risks and especially to those related to personal data collection and/or processing, adopt a risk-based approach, following the current regulatory trend, as provided for instance by the GDPR (Recital 75, 76) and AI Act proposal.

This approach requires to consider the risk of varying likelihood and severity for the rights and freedoms of natural persons. Following this approach, it is necessary to evaluate the ethics risks related to the data processing activities of the platform, assessing the likelihood and severity of each risk to data protection (or other ethical values), taking into account “the nature, scope, context and purposes of the processing and the sources of the risk”. The assessment of the risk must be conducted in an objective manner to determine whether there is a "risk" or a "high risk", to let the data controllers be particularly prudent to carefully consider their obligations when necessary. Such an approach requires consideration of what measures are appropriate in each case, depending on the scope, nature, context, and purposes of the processing concerned, as well as of the risks of varying likelihood and severity for freedoms and rights of individuals. The more severe and likely the risks from the proposed processing, the more measures will be required to counteract such risks. According to recital 75, examples of potentially risky processing relevant to a platform enabling the exchange of personal data include: i) processing that may give rise to discrimination, identity theft, financial loss, reputational damage, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage; ii) processing that might deprive data subjects of their rights and freedoms or prevent them from exercising control over their personal data; iii) processing of sensitive personal data; iv) processing for purposes of profiling; v) processing of personal data of vulnerable natural persons; Vi) processing involving a large amount of personal data and affecting a large number of data subjects.

For operationalizing the risk-based approach, the DataVaults project developed a specific tool and related service: the Risk Exposure Dashboard, displaying individual's current and projected risk estimations, which are updated whenever a modification to the shared assets occurs. Such estimations and risk exposure metrics are calculated relying on the data assets the data owner has already shared, as well as on the data he/she intends to share, and considering all sharing aspects, such as anonymisation level and discoverability, as well as the information provided by the nature of the data itself. The calculation of the privacy risk exposure, based on previous knowledge and depending on the data already available and shared and specific metrics, will allow also to notify individuals of their privacy risk exposure from the DataVaults Cloud Platform through the DataVaults Personal App.

The Risk Management Service might represent a high-value powerful accountability tool for the fulfilment of the GDPR-compliant informed consent requirement and user control, strengthening the positioning on the market of a data platform embedding it within its architecture and offering it to the individuals to foster their inclination to share their personal information.

A dedicated "Sharing Risk Information" operation is, indeed, essential for raising the awareness of the Individuals on the privacy exposure impact of sharing data assets.

6.1.1.3 Personas and Digital Twins

Depending on the defined data sharing configuration and selected user privacy level, different tools and techniques for privacy enhancement have to be used in a personal data sharing platform, ranging from the integration of traditional obfuscation schemes such as Digital Twins and User Personas, to the use of trusted computing technologies (i.e., TPMs) as a central element for the provision of privacy-preserving signature schemes based on the use of Direct Anonymous Attestation.

It is interesting here to elaborate on some legal and ethical challenges and opportunities raised by the Personas and the Digital Twins in relation to a personal data sharing platform or, in any case, to platforms based on personal data for their operation and service provision.

In DataVaults, the individuals can select the preferred level of anonymisation for the data asset they are going to upload and share in the DataVaults Cloud Platform: their personal data can be shared without applying anonymisation (eponymous), in an anonymized form or by sharing the data as a Persona utilizing the Persona Generator. More precisely, if the individual can select the anonymized sharing, depending on the individual's preference to upload and share as personal anonymized data (i.e., Digital Twin) or as grouped anonymized data (i.e., to become part of a Persona Group).

This is a very useful functionality related to the anonymization features and level of the personal data sharing, that should be available in any personal data platform.

As regards DataVaults, this is provided through the use of the Anonymisation Bundle, which, as regards Personas generation, groups data coming from different individuals and processes them using statistical methods for creating an aggregated representation/model where the individual's data is obfuscated by being included in a large pool of similar data, the so-called Persona.

In DataVaults, personas will be partially auto generated and presented to the data scientists prior to his/her analysis, based on certain similar aspects identified by the system (age group, location, interest, compensation requested, etc.). Through the DataVaults Cloud Platform the data scientist will be provided with an engine for the creation of aggregated profiles composed of data assets from several Individuals sharing certain similarities (generation of these Personas).

The creation of such personas is based on the obfuscation and merging of data originating from multiple users with similar characteristics: therefore, it is paramount to preserve their privacy. Such Personas are exactly aimed at preserving the privacy and anonymity of the distinct Individuals considered for the specific representation/model, though at the same time provide valuable information to data seekers. In DataVaults it is up to the individual to decide whether to share personal data in this way: the Individuals have indicated in the sharing configuration their intention to share data for the Personas generation. In any case, their privacy is protected, as all data assets to be shared under this condition, are appropriately anonymized prior to being transferred to the Cloud and being used in one or more Personas.

One of the challenges which need further investigation, in this regard, pertains to the revoked consent for data assets used for building Personas. All the data processing operations based on consent, which took place before the withdrawal, remain lawful but also that, in principle, any further processing of these data is prevented, if there is no other lawful basis justifying the continued retention and/or processing of the data. It is important to consider whether there is non-expired contract in place comprising such data assets: in that case, it is reasonable to conclude that the withdrawal can be exercised for the future without retroactive effect.

In case of withdrawal, the persona is still valid as insight from data even if the data provider withdraws consent – the data which was used can be deleted but the associated historical persona's may be retained– any update to the Persona would then not include data for which consent has been revoked. However, it is important to highlight that the insights cannot be attributed back to any one single individual so if they revoke consent their data is no longer being used anyway as any update to the persona would simply not include that individual's data within the aggregated dataset as it would no longer be available.

In relation to this issue, it is important to bear in mind different aspects: the individuals' right to withdraw consent anytime, the right to erasure/right to be forgotten and its boundaries (in consideration of the available technology, means and possible reasonable steps) and the other legitimate grounds for personal data processing and the limits to their applicability, with possible switching from one legal basis to another, as well as the interest of the data seekers. The legitimacy and fairness of technologies need to be sought by promoting the balance between competing interests and the determination of required level of protection for the personal information involved in these cases.

As regards the creation of Personas in the specific personal data platform concerned, it has also to be further explored if this implies or not some “profiling”, in the meaning provided by GDPR and therefore whether Art. 22 is applicable and, in case it is, if additional measures need to be taken. Persona’s do create profiles – in so far as people who meet criteria are included in the Persona so for example all people living in London, all people over 60 years old, all people with heart disease etc. etc. - the persona is then a much more granular profile of that group’s characteristics. However, this aspect goes beyond the DataVaults technology, since art. 22 refers to Automated Decision making. It would then be about the use of those Personas to make decisions such as who should be eligible for Health Insurance – this is outside of the scope of DataVaults – DataVaults does not control what the Data Seekers do with data they buy or persona’s they buy. In the post-project phase, it needs to be clarified on a case-by-case basis whether the use of the Persona is connected or not to an automated-decision making. It is important that the human intervention will be part of the task, especially in case some effects on the Individuals could occur (such as exclusion/limitation from some service or from a data sharing contract).

On the other hand, when the Data Scientists create the merged persona, the current user privacy risk exposure, as calculated by the DataVaults Risk Assessment framework, should respect the privacy choices defined by the user (in the data sharing configuration): in other words, the quantified privacy risk exposure values need to be kept within the user acceptable boundaries. Otherwise, the personal data platform should inform the user of appropriate actions to be taken for privacy enhancement.

Moving to the digital twins, first of all it is useful to provide a snapshot of that concept. “*A digital twin is a digital representation of a physical process, person, place, system or device”*. This concept, which emerged in the field of manufacturing domain, refers to digital simulation models that run alongside real-time processes [7] and it is conceptualized as digital replicas of physical entities, made possible using technological breakthroughs as sensing, processing, and data transmission. The digital twin concept is wide and can cover different aspects in different domains.

Within this overall debate on digital twins, this paragraph refers only to the personal digital twins relevant in the framework of a data platform based on personal data and their exchange using the elaboration of the digital replication of individual human data. We can refer to them as to Personal Digital Twins, since they reflect an individual (habits, history, behaviour, social interaction) and their personal data,

In particular, in the DataVaults project, the individual can configure the sharing anonymization level by selecting the preferred level of anonymisation for the data asset he/she is going to share, ranging from sharing data without applying anonymisation (eponymous), to anonymise personal data at an individual level (Digital Twin), or, as already mentioned, to anonymise them at a group level making them available for the creation of Personas. In case of selection of this data sharing configuration (Digital Twins), the DataVaults Cloud Platform shall generate the Digital Twin of an individual by anonymizing and obfuscating personally identifiable data while preserving the valuable information enclosed in the data asset, using the identity provided by another DataVaults component, the Identities Wallet. The Individual is allowed to view at any time under which Digital Twin Identities he/she has shared data anonymously with the DataVaults Cloud Platform.

Some of the ethical challenges potentially raised by the digital twins for instance based on data captured through Internet-of-Things-based sensing technologies have been initially explored by the narratives [8], thought the issues is still open. They might be potentially relevant during the future application of DataVaults, despite in an indirect way, since these aspects are outside the scope of DataVaults itself. In fact, the mentioned ethical challenges might become a concern when vendors collect private consumer information and are not transparent about how they use the data. Vendors must also protect that data. When organizations or bad actors can determine personal information from collected data and misuse it, IoT, ethics and privacy become intertwined. This is nor within scope of DataVaults as individuals sharing data have a responsibility to make sure IoT Vendors are transparent and gain consent etc.– as long as DataVaults gets the appropriate consent from the Individual, it has no control over anything else the vendor may or may not do with that data.

Of course, from an ethical point of view, there may be further aspects to consider concerning undesired side effects, despite they are not expected to be a consequence of DataVaults strictly speaking – as a secure privacy preserving data marketplace etc it transacts data and derivatives of that data – it plays no role in how that data and/or derivatives thereof are utilized.

Some of them are related to the nature of the human beings and the fact that people are often complex and adaptive to the changing environment: for instance, people can learn, exchange knowledge, have consciousness, are moved by goals changing over time, have emotions and so on. Similar characteristics might pose challenges for creating digital twins. There is the risk that application based on personal digital twins might interfere with individual thoughts, decisions and behaviours, human rights, and human dignity.

Other ethical concerns pertain, for example, on the risk of new forms of identity theft, abuse, and deception and how to mitigate them, as well as the risk that people are entirely replaced by digital twins.

A further concern regards the risk that personal digital twins are given greater opportunities and authority then human beings themselves, even though the digital representation of people and their desires could be biased, manipulated, or hacked. A personal digital twin and its data could be given more attention by the system/platform more than to humans, ignoring the opinion of the human the digital twin should represent.

There is also the risk of over- simplifications and of neglecting details and human dignity and other hardly measurable aspects, therefore overmining one of the main strengths of social systems: their ability to self-adaptation, self-organization and co-evolution, or, in other words, of a “technological determinism” of society.

To avoid that people could be managed like things and to prevent these risks to materialize, it is important to strongly rely on ethical mandates and soft law and the current regulatory reforms under development.

It is also critical that, rather than replacing individual preferences by automated machine decisions, to keep the individual’s control on their data and on the decision made relying on them. This is what systems like DataVaults are directed to, thereby minimizing the potential misuse of powerful digital technologies while maximizing benefits for the society.

6.1.1.4 Challenges related to Smart Contracts

To set, sustain and mobilize an ever-growing ecosystem for personal data and insights sharing and to foster an enhanced collaboration between individuals and data seekers, capable of rejuvenating the personal data value chain, an important aspect is to secure value flow based on smart contract safeguarding personal data ownership, privacy and usage and attributing value to the ones who produce it. Interesting approaches of personal data management make therefore use of smart contracts and DLT.

A major challenge for smart contracts however is how to transform them into contracts that preserve privacy and keep confidential information between the transaction parties, and in that manner disclose all relevant information from entities that do are not engaged in a transaction. In DataVaults, there was the need to go one step further, namely to conceal such information also from transaction parties, as it is essential in a personal data platform to provide clear and very high privacy guarantees to individuals providing personal data, as in many cases these users are comfortable to provide their data but at the same time do not want their data (or any transaction to be able to be linked back to them). For this purpose, the approach proposed and implemented in DataVaults was based on two dimensions:

a) enable a permissioned ledger deployment, so that no access is provided to external parties that are not part of the DataVaults platform
b) implement a double-ledger approach, where the DataVaults platform entity acts as a broker, and transactions are happening between the Data Owners and the platform, and between the Data Seekers and the platform, realising in this manner a transaction flow where Data Owners and Data Seekers are never part of the same transaction

For the purposes of a personal data sharing platform, it should be investigated if and how to ensure the electronic identification and to get the verifiable credential (on the basis of a national digital identity). For example, self-sovereign identity (SSI) and other identify management are relevant to attest ownership of data (to keep it sovereign etc).

The eIDAS Regulation states that the processing of personal data must be carried out in accordance with the GDPR and respecting its principle of confidentiality and security of processing: as clarified in its Recital 11, the authentication for an online service should concern processing of only those identification data that are adequate, relevant, and not excessive to grant access to that service online.

In case the platform concerned foresees to use electronic identification for its users, either natural or legal persons, this eIDAS Regulation can become applicable for its services and should be investigated especially in the context of the wallets and the smart contracts. Its electronic identification (eID) tools can be used for the identification of users, as they broadly offer enhanced security and accuracy, swifter, and less costly processes, while they may mitigate risk of fraud, identification theft and legal challenges.

On the other hand, the concept of self-sovereign identity [9] could also present advantages for the purpose of a personal data platform deployment and use and should therefore be investigated, including its compliance with eIDAS.

Sovrin argued that the "*self-sovereign identity (SSI) is a term used to describe the digital movement that recognizes an individual should own and control their identity without the intervention of administrative authorities. SSI allows people to interact in the digital world with the same freedom and capacity for trust as they do in the offline world*." Furthermore, "*Blockchain and SSI are natural complements, making the perfect symbiosis*": the user can individually create and manage his/her identify thanks to the use of distributed ledger technologies (e.g., blockchain), without the involvement of a third party, but often making use of the “decentralized identifier” (DID) associated with an entity. Such entity using SSI to authenticate itself can be an individual (natural person) and therefore, in this case, the DID usually relates to an identified or identifiable person (thus being personal data).

The SSI enables sovereignty for individuals over their digital assets and credentials, often by using digital wallets. In case the individual presents such assets and credentials to a third party to prove ownership, the public, decentralized, and immutable registry (such as a blockchain network) can be employed: the cryptographic proofs of the asset or credential were registered and are kept in a standardized and trustable way.

Nonetheless, the question whether eIDAS is already suitable for SSI and blockchain technology is still open, as well as whether, on the one hand, the smart contracts could be considered electronic documents and, on the other hand, the means used to sign blockchain transactions could be considered electronic signatures, with all the legal consequences it implies. Some scholars argue that the eIDAS Regulation will need some adjustments to become the legal and trust framework for SSI in the European Union: it was created as a legal framework supporting a digital identity metasystem mainly based in delegated authentication, which is more limited than the self-sovereign approach which enables, among other things, pseudonymity, and selective disclosure mechanisms.

In the US system the situation is not the same and some authors underlined that blockchain transactions can constitute, or evidence, electronic signatures, and that, virtually, all transactions stored on a blockchain, and retrievable in perceivable form, constitute an electronic record under the US law [10].

In conclusion, it is not fully clear whether for the purposes of operating a personal data sharing platform it should be ensured (and how) the electronic identification and it should be necessary to get the verifiable credential (based on a national digital identity), where necessary for accessing to online public services.

On the other hand, from the viewpoint of the smart contract itself, often used in the personal data platform to give the compensation for the sharing of own personal data, the debate is still ongoing whether and to what extent and conditions, these can give rise to legally binding and enforceable contracts and whether this necessarily requires the identification of the individual pursuant to eIDAS.

The smart contract satisfies the elements of a contract under several national laws, such as Spanish Civil Code and, therefore, smart contract code represents a valid mechanism to define the parties’ contractual rights and obligations as a matter of contract law in many jurisdictions. Therefore, “*under certain circumstances, and if so, decided by the parties, smart contracts can fulfil the elements of a legally binding contract under common law and civil law systems*” [11]. Though the parties may act pseudonymously, it is necessary a link (including off-chain) to their real identity to provide for valid consent, which is a crucial element of a contract under several national systems. However, even if its deployment does not give rise to a legally binding contract, the smart contract may still affect legal relations (either between the parties or with third parties) and therefore may have legal effects.

At the same time, both smart contracts and conventional natural language contracts can coexist in relation to the same (or related) subject matter and create together the entire legal framework within which a smart contract operates. This is the case of the so-called “external smart contract”, where “*the code does not form the entirety of the parties’ legal agreement,* *but merely automates the performance of some of its terms"*. The code merely automates the performance of some of the conventional contract's terms. In this case the legal relationship is intended to be governed by the natural language version of the contract, rather than by the code. In the internal model, on the contrary, the code could either encompass the entire agreement between the parties, or, alternatively, could form only an integral part of the legally binding contract (rather than the entirety of the contract), and would supersede any other clauses written in natural language: the code would be given legal effect and is an integral part of the agreement.

Principally, it is necessary to refer to the governing law applicable to the smart contracts to determine whether these give rise to legally binding contracts, whether personal identification is necessary or not according to eIDAS, as well as to evaluate the effects of the DTL/blockchain, and, ultimately, to ensure that the model chosen meets local law requirements. During the project, for the validation purposes, the user identities were verified manually. However, considering that in the future the DataVaults offering can constitute an electronic registered delivery service according to eIDAS (Art. 3, (36) eIDAS), such Regulations and the obligations established for the providers of such services have to be considered in relation to the future use of personal data platform.

6.1.2 DataVaults Ethical Policy, experiences and lessons learnt towards the effective personal data sharing under user control and benefitting all the actors involved

DataVaults is directed to rejuvenate the personal data value chain by delivering a framework and a platform having personal data, coming from diverse sources (wearables, web APIs, smart home sensors, personal data records, etc.) in its centre. Secure, trusted and privacy preserving mechanisms have been designed to allow the individuals to take ownership and control of their data and share them at will, through flexible data sharing solutions and fair compensation schemes with other entities (companies, public bodies, or other organisations). DataVaults aspires to become one of the flagship personal data platforms in the European landscape, characterized by fully respect of GDPR provision and satisfaction of the privacy and trust consideration of users, with a novel, fair and understandable value compensation mechanism to data owners.

Therefore, the consortium paid great attention to tackle any potential ethics issues raised by the platform's validation and future operation to give rise to a technology respectful of the data subjects' privacy and dignity and capable of prioritizing human well-being and flourishing. For this purpose, the Consortium elaborated the DataVaults Ethical Policy at the beginning of the project and adhered to it, conducted an in-depth regulatory review, elicited a set of legal and ethical requirements and related guidelines and recommendations for the overall DataVaults cloud-based platform and its components, as well as the Personal App and the demonstration activities. The Consortium also followed an Ethics and Data Protection Impact Assessment methodology, besides capturing the citizens' perspective through dedicated interactive channels.

This paragraph outlines the activities performed and outcomes achieved by the Consortium to adhere to the highest ethical standards and comply with the legislation, *in primis* the Data Protection Law (especially the GDPR).

6.1.2.1 DataVaults Ethical Policy, Approach and Ethical Requirements

The DataVaults Ethical Policy was conceived and implemented to ensure the legitimacy and fairness of project technologies and demonstration activities. It was elaborated at the beginning of the project, and it depicted the ethical procedures and responsibilities, including those relevant for human participation and personal data collection and processing in the demonstrators. In addition, it identified the oversight responsibilities (with the appointment and involvement in project's activities of the DataVaults Ethics & Data Protection Officer and DataVaults Ethical Board) and set the basis for the comprehensive Data Protection Impact Assessment methodology that was used during the demonstrators' operations. Such Policy also drew the roadmap for the implementation of ethics-related activities within the project.

The chosen Policy was driven by the Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals Approach. This approach was adopted for analysing the composite regulatory landscape, for deriving the legal and ethical requirements, as well as for providing recommendations and insights on how to face with the identified boundaries and constrains [12]. It was functional to ensure that the research activities, results, and validation activities are legally compliant and ethically sound. First, GDPR itself sets forth among the principles relating to processing of personal data the so-called "Fairness Principle". Fairness, which can be explained through the concepts of loyalty and good faith to be respected in all the steps of any personal data processing, requires that personal data must be used in a fair way, avoiding processing in a way that is unduly detrimental, unexpected, or misleading to the individuals concerned or that could have adverse impact on them. The "Fairness by Design" was considered as a straightforward requirement for DataVaults to ensure that individuals' privacy and real control over their data. The procedural dimension of the fairness entails the effective exercise of the data subjects' rights (rectification, erasure, object, etc.), whilst its substantive dimension implies moving towards the equal and just distribution of benefits and costs, without unfair bias, discrimination and stigmatization for individuals and groups.

This is linked with another high-level ethical requirement, the "sharing the wealth" paradigm [13]. It is aligned with the vision of a win-win data sharing ecosystem fostered by the DAIRO/ Big Data Value Association [2] as an effort to contribute to unlock the social value of personal data, going beyond user consent for fostering individual human empowerment and flourishing, as well as the common good of society and businesses' interests. The DataVaults Consortium followed this approach and directed its efforts to promote the alignment of its research and outcomes with social needs and expectations, also in view to strengthen the societal uptake of DataVaults cloud-based platform, given that high ethical standards generally imply public trust. This approach supports the identification, on a case-by-case basis, of the proper balance between competing interests and encompasses to societal fairness, based on equal opportunities and on the need to avoid that individuals are deceived or unjustifiably impaired in their freedom of choice. In view of fully ensuring the fairness of the technological artefact it is advisable to investigate several dimensions and consider different perspective, for instance focusing the attention on different kinds of compensation mechanisms, besides data monetization schemes, such as other rewarding incentives.

The chosen approach strongly relies also on human-centricity. Exploring and deepening individuals' viewpoint was considered essential by the Consortium for effectively adhering to the chosen Ethics, Fairness & Privacy-and-Security-by-Design-and-by-Default Approach and

[2] See, for instance, BDVA Position Paper "Towards a European Data Sharing Space - Enabling data exchange and unlocking AI potential" (April 2019).

for contributing to build a win-win data sharing ecosystem towards contributing to build public trust and, therefore, societal acceptance which is expected to enhance the uptake of data sharing technologies like DataVaults.

For this reason, the Consortium conducted a survey for capturing citizens' perspective, expectations, needs and concerns on personal data sharing: it was directed to individuals in their role of Data Owner. The survey was conducted online in 5 languages, and it comprised five groups of questions, respectively dedicated to:

- Attitudes towards personal data sharing
- Data retrieval, storage, and deletion
- Privacy preservation on the shared data
- Compensation Mechanisms
- Control and Informed Consent

The results from such survey, as well as of the other stakeholder engagement activities, were key for driving the design, development and deployment of the Personal Data Platform and App planned in DataVaults, whilst also providing important indications for the future progress of the Personal Data Market in Europe.

This attention to the individual is also consistent with the EC strategy and vision [14] directed to put people first in developing technology and to promote European values and rights in any design, development, and deployment of the technology in the real economy.

As regards the roles and responsibilities regarding ethics oversight, the DataVaults Ethics Board was set up and involved in the activities, as well as the Ethics and Data Protection Officer, who worked in collaboration with the Data Protection Officers or Ethics Responsible of the partners, especially the demonstrators.

The Board offered guidance, advice, monitoring and recommendations for future work, mainly with respect to ethics and privacy, whilst the Ethics and Data Protection Officer (EDPO) mainly supported the partners in ethics compliance and in the handling and management of personal data in accordance with the existing provisions of GDPR and other relevant EU and national legislations, providing guidance and advice, training of researchers, assisting in ethics risk assessment and supporting in relation to the Ethics and Data Protection Impact Assessments.

On the other hand, the Policy also drew the Ethical Procedures for the human involvement and personal data collection and handling in the demonstration activities, since individuals were involved in the pilots and their personal data, coming from diverse sources (sensors, IoT, wearables, data APIs, historical data, social network data, activity trackers, health records, demographic profiles, etc.) were gathered, processed, and shared. These procedures include those used to identify/recruit research participants, as well as the high-level description of the informed consent procedures for the participation of humans and personal data collection and processing, also including the sample of the informed consent/assent forms and information sheets distributed to the research participants. Such samples were fine-tuned and adapted by each relevant demonstrator, considering the specific context, technologies, and scenarios, with advice available from the EDPO where required.

The Ethical Policy guided the ethics-related work performed by the partners, both in the technical work-packages where the technological assets are designed and developed, and in the demonstrator phase, where they are assessed.

In particular, the Policy was strongly interrelated with the legal and ethical requirements elicitation. The Consortium in the early stage of the project set the Legal and Ethical Requirements for the design, development, and validation of DataVaults cloud-based platform and Personal App, as well as, to some extent, for the future operation of them, clearly laying out a first guideline for legal compliance and ethically-sound activities and results, without forgetting checkpoints. At a later stage of the project, the initial requirements list was extended considering the enriched legal review, where additional areas of law were analysed, as well as the regulatory reforms under development and their accompanying documents.

All these requirements were elicited adopting a systematic and holistic approach, driven by Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals method and relying on the Legal and Ethical Review. It consisted in the analysis of the regulatory landscape and the factual analysis of the privacy-relevant properties and personal data collection, processing and sharing in each service and tool, including details on the data categories, data sources and purposes of processing.

Some of the requirements were binding (when directly deriving from the legislation, such as GDPR), whilst others, where not directly imposed by the legislation, had to be interpreted more than recommendations or preferable requirements. Some requirements, being quite challenging, needed to be assessed with a certain degree of flexibility, considering the State-of-the-Art of the technological developments and the risk-based approach fostered by GDPR itself. In other words, this demanded for a certain degree of flexibility in the assessment of the adequateness of measures and technological solutions, to be specifically established on a case-by-case basis, considering a set of circumstances rotating around the severity of the risks and the reasonable efforts to face with them. It is recommended to adopt the same approach during the post-project phase, when DataVaults will be used in real-life environments.

The nature of the requirement was clearly stated in the description of each of them and they were provided in a table format for facilitating the quick understanding and reference to them by the technical team. Furthermore, in view of promoting the operationalization of the requirements, additional notes, recommendations, and guidelines were provided.

The fulfilment of the requirements regarding the DataVaults technology ensured that such technology adheres to legal and ethical mandates and is citizen centric. The Consortium conducted the assessment of the compliance with these requirements as part of WP6 evaluation activities in a triple iteration, respectively concerning the alpha, beta, and final version of the DataVaults technology.

6.1.2.2 The Ethics and Data Protection Impact Assessment Methodology

An important element of DataVaults Ethical Policy was the definition and implementation of the Ethics and Data Protection Impact Assessment methodology for the demonstrators, functional to the assessment of risks for individuals' rights, freedoms, and wellbeing, for ensuring compliance with the data protection law, and ethical mandates.

This methodology regarding the risks for the personal data was conducted following the indications of Article 35 sec. 1 GDPR, taking into account the nature, scope, context and purposes of the processing operations in each demonstrator in view of evaluating their impact on the protection of personal data, to identify and reduce the data protection risk [3] and the

[3] The concept of risk is clarified in Recitals 75-79 of the GDPR.

likelihood of privacy harms to individuals, as well as to identify and put in place the appropriate technical and organisational measures to tackle with/mitigate such risks.

The Consortium adopted a model inspired by the ISACA Model [15] for conducting such data protection assessment, which maps the fourteen ISACA privacy principles to the specific GDPR requirements and therefore allows an easy integration with any additional PIAs standards (Privacy Impact Assessment) required for other possible multiple privacy principles relevant for the demonstrators. Furthermore, this model is well aligned with the protection model focused on individual privacy and user control and efficaciously supports accountability, representing a useful instrument for the demonstrators to showing commitment and due diligence in taking adequate actions to ensure full compliance on an ongoing basis.

The DataVaults demonstrators (or pilots) elaborated their own EDPIA in conjunction with the respective technological supporting partners and the overall technical team of the project. It considered the specific DataVaults technologies (like services, components) relevant to their context, the data lifecycle, and each use cases scenarios, as well as their own privacy and security policies/practices.

Furthermore, for adequately also covering the ethical dimensions and for assessing to what extent the principle of fairness was operationalized in each of the demonstrator, the model inspired to the ISACA scheme was enriched with the Data Ethics Canvas [16]. This tool was elaborated by the ADAPT Centre for Digital Content Technology on the basis on the original Business Model Canvas by Alex Osterwalder. This model consists in a useful tool giving a higher-level framework to develop ethical guidance that suits any context and to assess the ethical implications of any project, thereby allowing to be more trustworthy with data processing. The Data Ethics Canvas can help those who collect, share and use data in identifying and managing ethical uses, both at the start of the initiative which imply data collection/processing and throughout [17]. On the other hand, thanks to it, the data seekers are supported in putting in place practices ensuring that the way the data is collected and used is trustworthy and ethical, beyond legal compliance. The Open Data Institute's Theory of Change is strongly consistent with the DataVaults' vision and with the Citizen Control of Personal Data Initiative within the Smart Cities Marketplace – Citizen Focus Action Cluster: "We want people who steward data, and people who create things with it, to act in ways that bring about positive impacts. Ethical use of data helps to improve trust and bring about the best economic and social outcomes. We want to avoid a future where data is feared or hoarded. We want data to work for everyone" [18].

As already mentioned, the Ethics and Data Protection Impact Assessments were conducted in each project's pilot through a questionnaire comprising elements coming both from the ISACA Model and from the Data Ethics Canvas. The EDPIAs in their consolidated release are reported in D6.5. Strong reference was made, besides to internal own policies, to the legal and ethical requirements set by the project itself. The EDPIA represented a key tool for ethical assessment and compliance in DataVaults and can be easily replicable, with the necessary adaptations, for use in other contexts for the future use of the DataVaults' innovation.

6.2 Lessons Learned from the Demonstrators' Experience

The demonstrators of DataVaults operated over a period of 22 months during the project (with an extra 6-month period as preparation activity), aimed to validate the technological outputs of the project as well as to validate the overall personal data sharing approach of the project, engaging individual users (Data Owners) as well as themselves (e.g., the personnel of the demonstrator organisations) as Data Seekers.

In principle, the different scenarios were run in 5 demonstrators which have been the following:

- Demonstrator #1 - Sports and Activity Personal Data, operated by Olympiacos Sports Club
- Demonstrator #2 - Strengthening Entrepreneurship and Mobility, operated by the City of Piraeus
- Demonstrator #3 - Secure Healthcare Data Retention and Sharing, operated by Andaman7
- Demonstrator #4 – Smart home Personal Energy Data, operated by MIWenergia demonstrator, as described in section 1.
- Demonstrator #5 - Personal Data for Municipal Services and the Tourism Industry, operated by City of Prato

More information about the different scenarios ran in the demonstrators and about the overall benefits and the evaluation of both the technical solution and of the business impact to the demonstrators is provided in deliverable D6.5 "Final Evaluation and Impact Assessment Report" which summarised the overall scenarios of the demonstrators and provides information relevant to the outcomes of those, the way they were executed and the benefits for the engaged target groups.

Combining the knowledge learned during both during the technical implementation activities, as well as the demonstrators' activities, a list of factors that are considered as crucial for the success of DataVaults (or any other personal data sharing platform) has been constructed.

Those factors, listed below, constitute a list that should be considered carefully by any organisation that wants to operate such a platform (therefore also technical success factors are provided), as well as for organisations that want to use such a platform as Data Seekers.

6.2.1 Critical Success Factors at Organisational Level

When it comes to organisational success factors that are considered as important for successfully operating an instance of the DataVaults platform (or a similar personal data sharing platform/infrastructure), those are the following:

- Clear Communication on Platforms' Concept. This is possibly the most important factor that needs to be considered by organisation that either operate such a platform, or that want to on-board Data Owners and other entities on the platform, as a clear communication strategy is required to convey to all interested stakeholders the high level concept of the platform, highlighting the security, privacy and ethics guarantees, and the benefits to be provided to all engaged stakeholder groups.

- Availability and Richness of Personal Data. As the driving force behind any data sharing platform is the availability of data, it is fundamental to work towards constantly increasing the availability and richness of personal data over the DataVaults platform. To do that, dedicated resources would be necessary to be allocated in activities for attracting both Data Owners and Data Seekers to the platform, as well as reaching agreements with other services that collect personal data for exposing them, and delivering new connectors for collecting individual's data from third party sources (see technical factors below)
- Meaningful Benefits and Incentives. As a major part of the personal data sharing concept (as implemented also in DataVaults) lies on the provision of benefits to Data Owners, it is essential to constantly look for and offer such benefits that will incentivise users to onboard the platform, and provide their personal data to Data Seekers, as they will get back some compensation for providing valuable personal information. This win-win relationship is not necessarily to be based on monetary (or tangible asset) compensation, and both Brokers and Data Seekers should identify benefits that really matter for Data Owners, and try to provide them, in a fair and responsible manner.
- Complete but non "boring"/ "discouraging" ethics forms. As practise has shown, Data Owners are not highly educated on issues relevant with privacy, security, and ethics (see below), and they tend to easily consent to everything. However, as ethics and privacy compliance are fundamental to DataVaults, such aspects are dealt in detail by the different components of the infrastructure and have a toll on user experience if these are not designed properly to put the minimal burden possible on end-users, so that they understand them and provide the required degree of informed consent.
- Simplification/explanation of main privacy terms. As stated above, users often complain about the complexity of the sharing process and the great amount of different privacy terms and options that they don't understand (e.g., anonymization, encryption, pseudo-ID, licenses, etc). In this context, it is essential to provide support to Data Owners to make them understand these terms, how these are related to their privacy and security exposure and to the ethical terms of the platform, and how these are handled by the different technological tools offered to them.
- Organisation Readiness and New Cost Centres Design. Aiming to operate the DataVaults as a "broker" requires certain changes in the organisation as identified in Section 4. Specific investment would be necessary and the definition of new roles and assignment of responsibilities within the organisation shall be done, as the operation of the platform can be seen as the introduction of a new business line (or activity) that shall be taken seriously as it bears a lot of responsibilities in terms of who to treat legal, business as well as reputation issues that may arise from its operation. Furthermore, new cost centres need to be designed, as this new business activity comes with specific costs that should be clearly considered when deciding to adopt and operate the platform as a broker.
- Employee Skills Cultivation. Finally, as the operation of the platform (either as a broker or a simple data seeker) requires the engagement of employees with certain skills, it is necessary for the organisation to invest in educating employees on aspects such as data management and sharing, as well as in security and privacy concepts. Moreover, skills on data analytics are also considered very important for both brokers and data

seekers, as these would allow organisations to exploit the real power of the datasets acquired and increase the added value that the platform can provide to them.

6.2.2 Critical Success Factors at Technical Implementation Level

Regarding Technical Success Factors, these refer to factors that should be considered by organisations that are interested to operate the technical infrastructure of DataVaults, either as brokers, or for satisfying their own needs (to connect with their client base). The things to consider in this dimension are the following:

- Design an intuitive UI and provided an easy and attractive UX. It is very important to have a nice and effective interface, with limited number of steps for each procedure, particularly for Data Owners who may use their smartphone or their PC and who may not be skilled enough to follow complicated paths. Moreover, having an attractive and easy to use interface motivates users to spend more time on the platform as their degree of satisfaction is rather high, refraining in this case from user dropouts.
- Provide ready-made templates to hide the complexity of data sharing. As identified above, the more powerful data sharing features are offered to users, and the more sophisticated tools become, the more the complexity of the overall workflow is increasing. It is therefore important for operators to design smart ways to hide this complexity (not in expense of security and privacy) and deliver to the users different interfaces that allow them to quickly perform their tasks, at least the ones that are considered mainstream or repetitive.
- Provide high and easy to understand/select security & privacy guarantees. In conjunction with the above-mentioned factor, it is crucial that the guarantees provided to Data Owners (As well as to Data Seekers) are of the highest degree and at the same moment allow them to use them or select any other level without investing a lot of effort in such activities. This can be achieved through the offering of ready-made filled-in templates which can include pre-selected options regarding the different privacy and security levels alongside with fast to digest contextual information, to allow users to take the right decisions, in a very comfortable, trusted, and easy manner.
- Dedicate Resources to increase supported of Data Sources. As the availability and the richness of data over the platform constitutes the major success factor, it is very important for organisations to dedicate IT resources to continuously support, extend these capabilities of the platform by working on the maintenance and the development of new connectors to third-party data sources and on other ways that can be used by Data Owners to collect their data into the platform, and make them at a later stage available for sharing.
- Deploy the platform over robust and performant infrastructure. Another critical factor, as system performance drastically affects user experience and having a slow performing system will demotivate users (mostly Data Owners) from using it.

Whilst this section covers the direct experiences of the demonstrators involved directly in the deployment of the platform, a host of other lessons learned have been gathered by partners and published in the DataVaults led book: “Personal Data Smart Cities”, looking at the project in a more holistic way [5].

7 CONCLUSIONS

The present deliverable outlines the knowledge distilled from the DataVaults demonstration activities relevant to best practises and information that should be made available to potential adopters of the DataVaults solution, to let them understand how they should work toward scaling up the deployment they wish to operate. Furthermore, some of the findings during the demonstrator activities, have been extended to provide a helping hand to organisations that are considering other similar solutions that are based on the same principles (e.g. personal data sharing), providing them the DataVaults perspective of the main strategies one has to devise in order to achieve its goals in this domain.

As identified in this deliverable, there are specific pre-requisites that need to be covered by such organisations, that deal with operational, technical, and legal readiness and compliance. Such pre-requisites have to be in place prior the operation of the platform.

Furthermore, a set of lessons learned from the legal point of view have been presented, alongside with experiences (captured as benefits to stakeholder groups) recorded by the DataVaults piloting partners. They clearly reveal the way that such a platform can provide added value to the whole data sharing ecosystem in general, and also to the different target groups, showcasing the value of personal data, as well as the need to tackle it in a responsible and thorough way, adopting certain methodologies for the management of the legal and ethical issues that usually surround these data.

Concluding, a set of critical success factors are presented, which are expected to be of outmost importance for any potential adopter, highlighting the main lines along which an organisation that would like to operate and use DataVaults (or a similar personal data sharing platform) should operate and plan its strategy.

REFERENCES

[1] https://jupyter.org/

[2] https://mlflow.org/

[3] https://superset.apache.org/

[4] M. Da Bormida in Cugurra (2022). Does Everything Conform to Legal, Ethical, and Data Protection Principles. Available at: https://www.riverpublishers.com/pdf/ebook/chapter/RP_9788770227995C15.pdf

[5] https://www.riverpublishers.com/research_details.php?book_id=1033

[6] Big Data Value Association (BDVA), "Data protection in the era of Artificial Intelligence", 2019.

[7] Grieves M. Grieves M. (2014). Digital Twin: manufacturing excellence through virtual factory replication. White Paper

[8] D. Helbing, J.A. Sanchez-Vaquerizo (2022). Digital twins: Potentials, Limitations and Ethical Challenges

[9] M. Allende Lopez (2020). Self-sovreign identity. The future of Identity: self-sovreignity, Digital Wallet, Blockchain

[10] Notably the Electronic Signatures in Global and National Commerce Act (2000). ESIGN: public law, pp. 106-229

[11] Smart Contract Alliance (2018). Smart Contracts: is the Law Ready?

[12] DataVaults D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management".

[13] M. Da Bormida (2021). The Big Data World: benefits, threats and ethical challenges" in Advances in Research Ethics and Integrity

[14] "A European strategy for data" COM (2020) 66 final.

[15] ISACA (2017). GDPR Data Protection Impact Assessment7

[16] W. Reijers, K. Koidl, D.L. Harshvardhan J. Pandit & B. Gordijn (2018). Discussing Ethical Impacts in Research and Innovation: The Ethics Canvas. Part of the IFIP Advances in Information and Communication Technology book series)

[17] Open Data Initiative (ODI) (2017). Helping organizations navigate ethical concerns in their data practices.

[18] https://theodi.org/theory-of-change