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

D1.3
DataVaults MVP and Usage Scenarios

Editor(s)	Sotiris Koussouris, Stefanos Venios, Nefeli Bountouni, Marios Phinikettos
Lead Beneficiary	Suite5
Status	Final
Version	1.00
Due Date	30/09/2020
Delivery Date	06/10/2020
Dissemination Level	PU

This deliverable has been submitted to the EC and is pending approval

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

Project	DataVaults – 871755
Work Package	WP1 - DataVaults Data Value Chain Elaboration and Concept Fusion
Deliverable	D1.3 - DataVaults MVP and Usage Scenarios
Editor(s)	Suite5 - Sotiris Koussouris, Stefanos Venios, Nefeli Bountouni, Marios Phinikettos
Contributor(s)	ANDAMAN7 – Sebastien Hanney Assentian – Shaun Topham and Ilesh Dattani from ATOS – Ricardo Ruiz, Javier Villazán and Miriam Quintero ETA - Marina Da Bormida FOKUS – Kyriakos Stefanidis IFAT – Alexander Köberl MAGGIOLI - Nikos Achilleopoulos, George Boukis, Andrea Montefiori, MIWenergia – Pablo Barrachina, Ramón Ruiz OLYMPIAKOS – George Totomis, Christina Tsiligkiri Paolo Mattarelli Prato – Paolo Boscolo, Elena Palmisano UBITECH – Giannis Ledakis Unisystems – John Kaldis
Reviewer(s)	IFAT – Alexander Köberl UBITECH – Giannis Ledakis

Abstract	This deliverable defines the integrated DataVaults Methodology and develops a set of high-level operation scenarios that help identify the expected behaviour of the DataVaults platform, leading to the formulation of the DataVaults Most Valuable Product (MVP).
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 D1.3 - "DataVaults MVP and Usage Scenarios" documents the work performed in the context of T1.5 - "DataVaults MVP Definition and High-Level usage Scenarios" and presents the produced results. The purpose of this task is to define the integrated DataVaults Methodology and develop a set of high-level operation scenarios that help identify the expected behaviour of the DataVaults platform and lead to the formulation of the DataVaults Most Valuable Product (MVP). The outputs of this task reveal the workflow and interrelations between the different concepts of DataVaults, while shedding light on the different stakeholders and their expectations from the platform.

The initial DataVaults Methodology is elaborated in this deliverable. It is structured in **eight** interrelated Phases, aiming to capture the scope of DataVaults. Each Phase has its own Operations that outline at a high level the distinctive tasks that will be performed by the platform. More specifically, the DataVaults Methodology comprises the following Phases and Operations:

Phase I: Data Retrieval

- Operation I.1: Data Retrieval Configuration
- Operation I.2: Data Retrieval Implementation
- Operation I.3: Data Asset Retrieval Authorisation Revocation

Phase II: Data Transformation & Enrichment

- Operation II.1: Data Quality Check & Transformation
- Operation II.2: Semantic Enrichment & Annotation
- Operation II.3: Data Linking

Phase III: Asset Storage at Individual's Side

- Operation III.1: Asset Generation and Storage
- Operation III.2: Encryption at the Personal DataVaults Side
- Operation III.3: Local Asset Storage
- Operation III.4: Local Asset Indexing
- Operation III.5: Local Asset Deletion

Phase IV: Asset Sharing to DataVaults Cloud Platform

- Operation IV.1: Data Sharing Configuration
- Operation IV.2: Sharing Risk Information
- Operation IV.3: Private Contract Management - Individual & DataVaults Cloud Platform
- Operation IV.4: Compensation Management

Phase V: DataVaults Cloud Platform Asset Storage

- Operation V.1: Encryption and Access Policies Enforcement
- Operation V.2: DataVaults Cloud Platform Asset Storage
- Operation V.3: DataVaults Cloud Platform Asset Indexing
- Operation V.4: DataVaults Cloud Platform Asset Deletion

Phase VI: Asset Exploration & Extraction

- Operation VI.1: Asset Searching
- Operation VI.2: Public Contract Management - Data Seeker – DataVaults Platform
- Operation VI.3: Asset Export

Phase VII: Data Analytics

- Operation VII.1: Cloud Based Data Analysis
- Operation VII.2: Data Visualisation
- Operation VII.3: Persona Creation

Phase VIII: Added Value Services

- Operation VIII.1: Notifications
- Operation VIII.2: Sharing Gains

Afterwards, the **five main Actors** within the DataVaults ecosystem are defined, to better understand their incentives and foresee their interactions. The identified actors are: the Individuals, the Data Seekers, the DataVaults Data Scientist and finally the DataVaults platform itself, distinguished in the DataVaults Personal App and the DataVaults Cloud Platform. **Eleven** high level platform operation scenarios are outlined, based on the demonstrator scenarios and gathered feedback, to represent the different needs of users and guide through the envisioned functionalities. The scenarios are divided in three groups, around the main involved Actors (i.e. driven by the Individual, the Data Seeker, or the DataVaults Data Scientist). They include not only the scenario descriptions, sequence of steps and workflow diagrams, but investigate ethical, security and GDPR-related aspects, in order to ensure compliance, highlight points that shall be researched in depth and pinpoint functionalities that are required for achieving maximum legal compliance, privacy and security.

Finally, the present deliverable delineates the first version of the DataVaults MVP. The first step towards the construction of the MVP, was the extraction of a **set of 79 features** from the integrated methodology and the operation scenarios. These features are distinguished to Platform and Personal, based on whether they are envisioned for the DataVaults Personal App or the Cloud Platform respectively. The features were also mapped to the related Methodology Phases at Operation level, as well as to the Scenarios, for a better view of how everything will come together. Subsequently, the homogenised features were evaluated from a technical and business aspect by the partners, to accommodate a first prioritisation. The demonstrators were involved in the business value assessment, using the MoSCoW (Must-have, Should-have, Could-have, Won't-have) voting technique [1], while technical partners evaluated the features on their value for the platform as well as their complexity, using a two-factor voting process. The voting results were processed and correlated, to end up with the initial DataVaults MVP, consisting the features categorised in four groups, based on the complexity matrix results, and labelled with their added business value, thus providing guidance for their prioritization in the implementation tasks. The results of this deliverable, including the integrated methodology, operation scenarios and MVP, will be leveraged for the elicitation of the DataVaults user stories and requirements, in WP3, WP4 and WP5. As such, part of this deliverable will be revised, as planned in the DOA, based on the final DataVaults Architecture, the requirements and the outcomes of the platform development, and will be documented in the context of D1.4 in M18 of the project.

Table of Contents

1 Introduction 10
1.1 Document Approach 10
1.2 Relation to Other WPs and Deliverables 11
1.3 Document Structure 11
2 The DataVaults Methodology 12
2.1 Phase I: Data Retrieval 15
2.1.1 Operation I.1: Data Retrieval Configuration 15
2.1.2 Operation I.2: Data Retrieval Implementation 15
2.1.3 Operation I.3: Data Asset Retrieval Authorisation Revocation 15
2.2 Phase II: Data Transformation & Enrichment 16
2.2.1 Operation II.1: Data Quality Check & Transformation 16
2.2.2 Operation II.2: Semantic Enrichment & Annotation 17
2.2.3 Operation II.3: Data Linking 17
2.3 Phase III: Asset Storage at Individual's Side (DataVaults Personal App) 17
2.3.1 Operation III.1: Asset Generation and Storage 18
2.3.2 Operation III.2: Encryption at the Personal DataVaults Side 18
2.3.3 Operation III.3: Local Asset Storage 19
2.3.4 Operation III.4: Local Asset Indexing 19
2.3.5 Operation III.5: Local Asset Deletion 19
2.4 Phase IV: Asset Sharing to DataVaults Cloud Platform 20
2.4.1 Operation IV.1: Data Sharing Configuration 20
2.4.2 Operation IV.2: Sharing Risk Information 21
2.4.3 Operation IV.3: Private Contract Management – Individual & DataVaults Cloud Platform 21
2.4.4 Operation IV.4: Compensation Management 22
2.5 Phase V: DataVaults Cloud Platform Asset Storage 22
2.5.1 Operation V.1: Encryption and Access Policies Enforcement 22
2.5.2 Operation V.2 DataVaults Cloud Platform Asset Storage 23
2.5.3 Operation V.3: DataVaults Cloud Platform Asset Indexing 24
2.5.4 Operation V.4: DataVaults Cloud Platform Asset Deletion 24
2.6 Phase VI: Asset Exploration & Extraction 24
2.6.1 Operation VI.1: Asset Searching 24

2.6.2 Operation VI.2: Public Contract Management - Data Seeker – DataVaults Platform 25
2.6.3 Operation VI.3: Asset Export 25
2.7 Phase VII: Data Analytics 25
2.7.1 Operation VII.1: Cloud Based Data Analysis 25
2.7.2 Operation VII.2: Data Visualisation 28
2.7.3 Operation VII.3: Persona Creation 28
2.8 Phase VIII: Added Value Services 28
2.8.1 Operation VIII.1: Notifications 28
2.8.2 Operation VIII.2: Sharing Gains 28
3 DataVaults High Level Platform Operation Scenarios 29
3.1 Actors Definition 29
3.1.1 Individual 29
3.1.2 Data Seeker 29
3.1.3 DataVaults Personal App 30
3.1.4 DataVaults Cloud Platform 30
3.1.5 DataVaults Data Scientist 30
3.2 Scenarios driven by Individuals 30
3.2.1 Scenario 1: Personal Data Collection 30
3.2.2 Scenario 2: Personal Data Assets Exploration & Analysis 34
3.2.3 Scenario 3: Personal Data Assets Sharing Gains and Risk Information 37
3.2.4 Scenario 4: Personal Data Assets Sharing Configuration 39
3.2.5 Scenario 5: Personal Data Sharing / Cloud Upload 43
3.2.6 Scenario 6: Personal Data Assets Sharing Revocation and Deletion 45
3.3 Scenarios driven by Data Seekers 48
3.3.1 Scenario 7: Explore Data Assets 48
3.3.2 Scenario 8: Acquire Data Assets from the DataVaults Cloud Platform 50
3.3.3 Scenario 9: Acquire Data Assets from a DataVaults Individual User 52
3.3.4 Scenario 10: Analyse and Visualise Data 55
3.4 Scenarios driven by the DataVaults Data Scientist 57
3.4.1 Scenario 11: Ready-Made Analysis by the DataVaults Cloud Platform 57
4 Features Extraction 60
5 The DataVaults MVP – Version #1 80
5.1 What is an MVP 80

5.2 Features Value Assessment 81
5.3 MVP Consolidation – Version #1 85
6 Conclusions and Next Steps 88
7 References 89
ANNEX I: Complete Results of Feature Voting 90
ANNEX II: Demonstration Scenarios – Operation Scenarios Mapping 101

List of Tables

Table 1 - Toreador Algorithms 26
Table 2 - Reverse Mapping of Average Feature Scores to MoSCoW Labels 82
Table 3 - Highly Valuable and Not Complex 85
Table 4 - Highly Valuable and Complex 86
Table 5 - Not Highly Valuable and Not Complex 87
Table 6 - Not Highly Valuable and Complex 87

List of Figures

Figure 1 - Document Approach 10
Figure 2 – The DataVaults Methodology and its Relation to the DataVaults Data Lifecycle ... 14
Figure 3 - DataVaults MVP Approach 81
Figure 4 - Visualisation of Feature Voting Results 84

Terms and Abbreviations

API	Application Programming Interface
BDVA	Big Data Value Association
DAA	Direct Anonymous Attestation
DoA	Description of Action
Dx.y	Deliverable x.y
GDPR	General Data Protection Regulation
MoSCoW	Must-have, Should-have Could-have, Won't-have
MVP	Most Valuable Product
SS	Scenario Sequence
Tx.y	Task x.y
WPx	Workpackage x

1 INTRODUCTION

Deliverable D1.3 aims at developing the high-level usage scenarios and formulating the DataVaults Most Valuable Product (MVP). More precisely, its purpose is to reveal how concepts interrelate and, starting from the project's pilot partners and the extended network of the consortium, to indicate all stakeholders' point of view and capture the most important needs of the users. The work in D1.3 is part of WP1 and more specifically, in the context of task T1.5 "DataVaults MVP Definition and High-Level usage Scenarios".

One of the major outcomes of T1.5 is the DataVaults Methodology. A detailed analysis of the as-is and to-be processes is performed, illustrated as workflow diagrams for different prospective platform users, revealing the logical flow of information and operations both in the DataVaults platform as well as the DataVaults Personal App. This task is based on the definition of high-level usage scenarios. The methodology defined leads towards the formulation of the DataVaults MVP, which covers the most important needs of the users and prioritizes the features to be transferred into implementation and deployment.

1.1 DOCUMENT APPROACH

The deliverable follows a clear and comprehensive approach in order to derive the outcomes of T1.5. Figure 1-1 depicts a high-level and abstract overview of the approach followed.

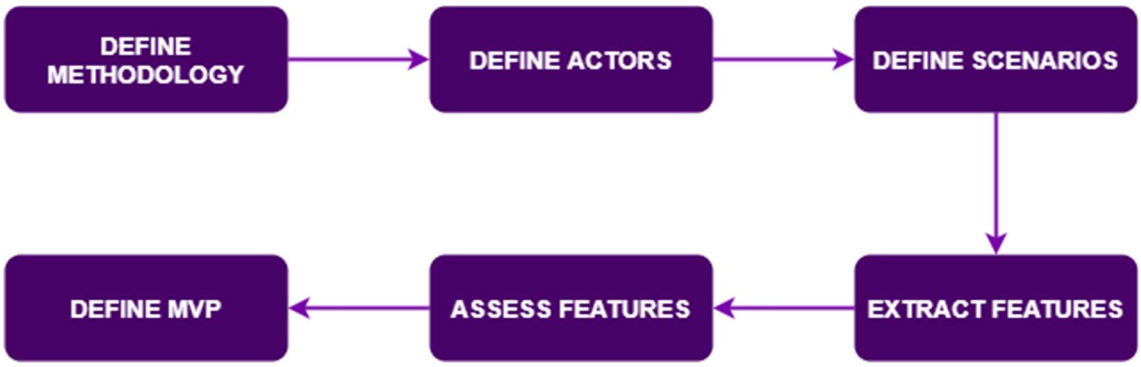


Figure 1 - Document Approach

At first, the key findings of D1.1 "DataVaults Data Value Chain Definition" have offered valuable input to the definition of the DataVaults Methodology. The methodology is divided in various phases, with each phase having its own specific operations and aspects (as detailed in section 2).

As a second step, the five DataVaults Actors that are involved in the interactions within the DataVaults ecosystem, were identified. Afterwards, the consortium defined several high-level usage scenarios of DataVaults based on the DataVaults methodology, DataVaults demonstrators and key findings from D1.1 "DataVaults Data Value Chain Definition". Eleven high-level scenarios were defined in detail, as representative scenarios of all DataVaults actors (as detailed in section 3).

The derived methodology along with the possible high-level scenarios and the demonstrators' requirements have facilitated the extraction of features that could be possibly included in the

DataVaults MVP (presented in section 4). Considering these features, a voting process among the members of the consortium was organised in order to specify feature importance and lead to the selection of the ones that constitute the first version MVP (presented in section 5) that will be used to drive the initial development tasks of the project.

1.2 RELATION TO OTHER WPS AND DELIVERABLES

This deliverable is the outcome of the T1.5 "DataVaults MVP Definition and High-Level usage Scenarios" which remains active until the 18th month of the project. It used D1.1 "DataVaults Data Value Chain Definition" taking input from the demonstrator scenarios and problem statements, to elicit the common DataVaults usage scenarios and define the DataVaults. Deliverable D1.2 "The DataVaults Core Semantic Data Model" provided the DataVaults Lifecycle, that was the basis for the integrated DataVaults Methodology.

The DataVaults Methodology, revealing how components and concepts interrelate and displaying high level usage scenarios of the concept, formulating the platform's MVP, will provide input to the use cases, the architecture and specification tasks in WP3, WP4, WP5. It will also support the marketing end exploitation activities in WP7.

1.3 DOCUMENT STRUCTURE

The following sections of the specific deliverable are structured as follows:

- Section 2 presents the definition of the DataVaults Methodology and is divided in eight phases (Phase I - Data Retrieval, Phase II - Data Transformation & Enrichment, Phase III - Asset Storage at Individual's Side, Phase IV - Asset Sharing to DataVaults Cloud Platform, Phase V - DataVaults Cloud Platform Asset Storage, Phase VI - Asset Exploration & Extraction, Phase VII - Data Analytics, Phase VIII – Added Value Services), with each phase having its own specific steps;
- Section 3 presents the defined high-level usage scenarios of DataVaults for different prospected users, that present the logical flow of information and operations both in the DataVaults platform as well as the DataVaults Personal App;
- Section 4 presents the platform features that are extracted from the methodology and are related to the high-level scenarios;
- Section 5 presents the initial internal assessment of the features and the preliminary consolidation of the DataVaults MVP;
- Section 6 concludes this deliverable.

2 THE DATAVAULTS METHODOLOGY

This section presents the methodology behind the high-level concept of DataVaults. Part of this methodology will drive the DataVaults platform development, by providing a comprehensive way to better understand what services shall be offered by the Platform and how they relate to each other. The DataVaults Methodology is constructed in eight interconnected Phases, aiming to capture the overall scope of the project, by outlining at a high-level, any interactions that may occur between the DataVaults Actors and the DataVaults Platform. Horizontal services were also identified, to act supportively to all Phases.

The Phases stem from the DataVaults Lifecycle, that has been outlined in the context of deliverable D1.2 "The DataVaults Core Semantic Data Model". In particular, the overall DataVaults Lifecycle has been broken into three parts corresponding to the three workflows that have been delineated as the core offerings of the Platform:

- **Data Management Lifecycle**: This part of the Lifecycle concerns all data management-related tasks, spanning from data collection, data cleansing and semantic enrichment, storage and sharing, up to data deletion and access revocation.
- **Data Analytics Lifecycle**: It contains the steps required for the exploration of data assets and the application of data analytics and visualisation techniques in order to extract meaningful insights.
- **Compensation Lifecycle**: This part of the Lifecycle evolves around the creation and management of contracts for the sharing of data and the appropriate compensation of Individuals.

Each Phase of the DataVaults Methodology is dedicated to a specific task and has been designed to group closely related Operations, thus facilitating a modular development approach.

A brief overview of the eight Phases of the Methodology is provided below:

i. **Phase I: Data Retrieval** – refers to the configuration, implementation and management of the DataVaults connection to the various data sources, as defined by the Individuals, in order to collect their personal data.
ii. **Phase II: Data Transformation & Enrichment** - ensures the high quality of the collected data, through automated quality checks and transformation operations. Furthermore, data schema mapping, semantic enrichment and linking processes are foreseen for the maximisation of discoverability and usability of these data.
iii. **Phase III: Asset Storage at Individual's Side** – is responsible for the persistence of the personal data assets at the DataVaults Personal App of the Individuals in a secure way. These assets include the processed and semantically enriched data collected from the connected data sources, their metadata, and any data assets generated from the application of data analytics by the individual.
iv. **Phase IV: Asset Sharing to DataVaults Cloud Platform** – entails all aspects around the sharing of an Individual's data asset to the DataVaults Cloud Platform. These processes span from the configuration of the various sharing aspects and the creation and management of the corresponding contracts between the Individuals and the

Platform, to the continuous update of the Individual's sharing risk information and the activation of the compensation mechanism whenever a data asset is acquired by a Data Seeker.

v. **Phase V: DataVaults Cloud Platform Asset Storage** – pertains to the actual upload and secure storage of a data asset from the Individual's side to the Cloud Platform, under the terms set during the sharing configuration. Furthermore, this Phase handles a requested deletion of the data asset from the Cloud.

vi. **Phase VI: Asset Exploration & Extraction** – enables the Data Seekers search and acquire data assets that match their needs.

vii. **Phase VII: Data Analytics** – provides cloud-based data analytics and visualisation tools that will shed light on underlying connections and facilitate Data Seekers into getting a better understanding.

viii. **Phase VIII: Added Value Services** –includes horizontal DataVaults services that facilitate other core processes.

Although the Phases are numbered, their execution is not consecutive. On the contrary, only some of the Phases and Operations are activated during the execution of the various use cases. Any interdependencies, such as an Operation or Phase being a prerequisite for others, have been considered and are denoted within the Features that have been derived from the Methodology and Scenarios, in the following sections. It should be noted that user and profile management operations, although will be implemented as part of the DataVaults Platform, are not included the DataVaults Methodology, as the main focus is on data related tasks. The following Figure 2, demonstrates the interrelations of the Methodology at Phase level, as well as their 'mapping' to the DataVaults Data Lifecycle that was presented in D1.2:

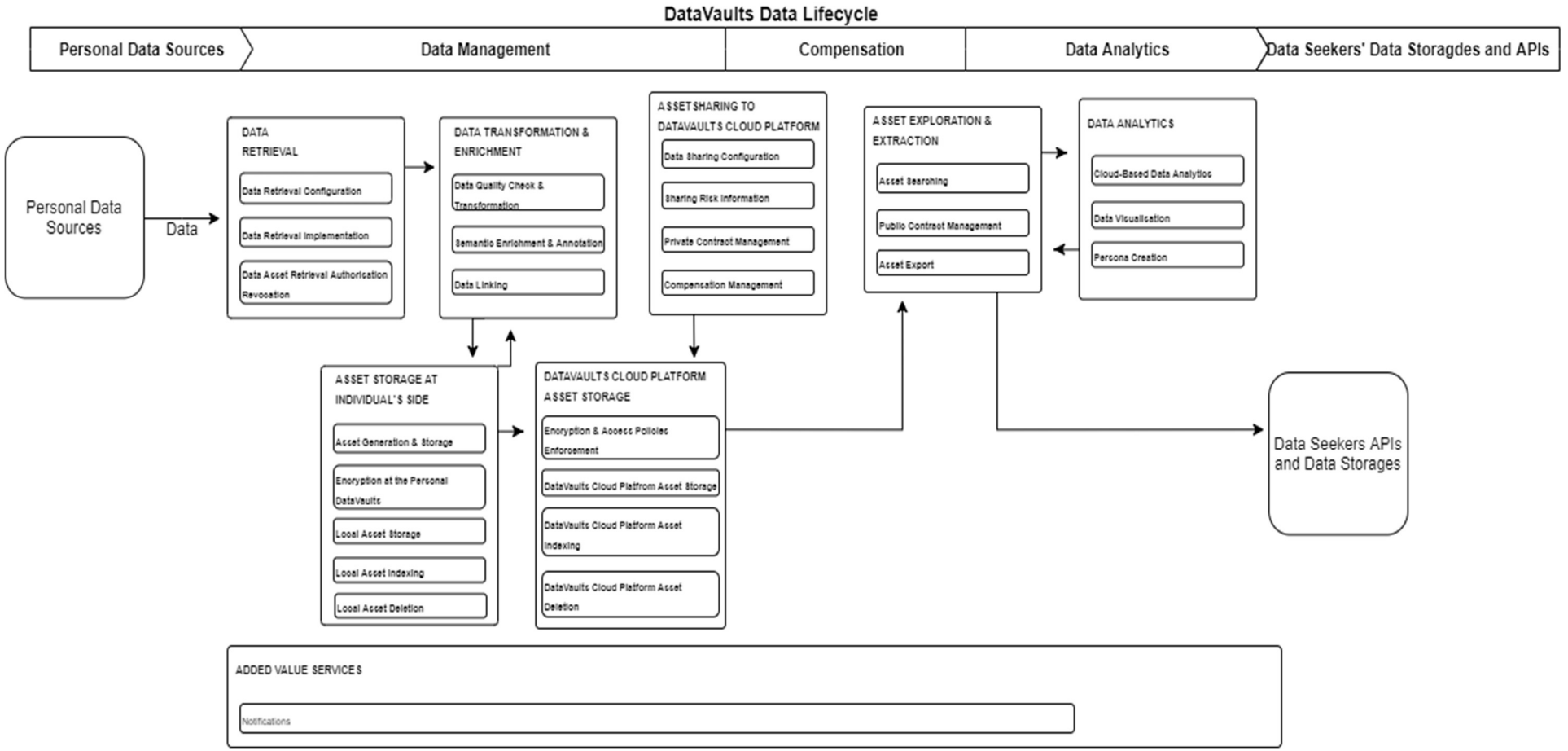


Figure 2 – The DataVaults Methodology and its Relation to the DataVaults Data Lifecycle

2.1 PHASE I: DATA RETRIEVAL

This Phase is dedicated to all actions revolving around the collection of personal data from data sources indicated by the Individuals. It is associated not only to the configuration and establishment of the required connections from DataVaults Personal App to the data sources, but also to enabling the user to revoke retrieval authorization in case he/she does no longer wish any more data to be collected from a specific source.

2.1.1 Operation I.1: Data Retrieval Configuration

Data Retrieval Configuration constitutes the initial operation of the Data Retrieval phase. This operation is responsible for creating the required technical bridges to acquire access to the personal data of the Individual. These personal data are either to be provided manually by the Individual (i.e. demographics, personal preferences, replies to ad-hoc questionnaire, etc.) or will be collected by the DataVaults app in a semi-automatic manner from third-party apps and data sources (i.e. wearables, smart home devices, other smartphone apps installed on the device, etc.) through available APIs or through direct access authorization.

For the semi-automatic Data Retrieval through an API, the Individual is prompted during the Data Retrieval Configuration to provide the required user credentials that will be utilised as tokens for the establishment of the API connection, or to authorize DataVaults in order to gain access to the data stored by the third-party app/source.

Then, the Individual is provided with additional options relevant to the authorisation process and data retention, e.g. data retrieval access/user credentials expiration date, interval of data collection, frequency of data collections, how should obsolete/expired data be handled (for example, with auto-deletion), auto-renewal of the operation upon token/data asset expiry, etc.

2.1.2 Operation I.2: Data Retrieval Implementation

Data Retrieval Implementation corresponds to the actual connection of the Personal DataVaults App with the data source, based on the configuration compiled previously. This operation is crucial in ensuring the successful implementation and deployment of the APIs and access authorizations that have been configured in the previous operation.

After the establishment of the connection, sample dataset(s) are gathered from the data source to verify that the requested data can be actually retrieved through the provided gateway.

2.1.3 Operation I.3: Data Asset Retrieval Authorisation Revocation

This operation concerns the case where the Individual has decided to revoke the authorization previously provided to the DataVaults Personal App to collect personal data from data sources. Such personal data are assumed to have been collected by the DataVaults Personal App and stored at the user's side but not yet configured for sharing, shared and uploaded to DataVaults Cloud. This revocation could be for one or more data sources. The Individual de-authorises the DataVaults Personal App to collect personal data by forcing the de-registration/deletion of active tokens, thus the DataVaults app is no longer able to connect to the indicated source and collect data.

2.2 PHASE II: DATA TRANSFORMATION & ENRICHMENT

The Data Transformation & Enrichment Phase is related to the automated processing of the raw data that have been collected from the data sources, in order to ensure high quality and reliability, as well as to maximise their semantic value and discoverability through their compliance and mapping to the pre-defined DataVaults schemas.

2.2.1 Operation II.1: Data Quality Check & Transformation

The datasets ingested during the data retrieval process, do not always come in the expected form to be further utilised in DataVaults, and may even contain erroneous or unnecessary data. In order to ensure the integrity and reliability of the collected datasets, an appropriate Data Quality Check process is designed, aiming to discover inconsistencies and other anomalies in the data through validation rules and various statistical checks. Following this, a Data Transformation procedure will take place. Data Transformation is highly connected with the results of the Data Quality Check, as these results will point out the cleaning and transformation techniques that shall be applied per case. These techniques include data replace, reduce, data normalization, and other data wrangling methods.

Initially, after the data sample has been collected through the established connection, an automatic process is initiated, in order to assess the data quality, including the particular standards that the data must conform to. This process is facilitated by a set of validation rules that are defined by the DataVaults Data Scientist. Various aspects of data can be assessed with these rules, including among others: verification of the dataset conformity to certain standards, ensuring that the data have no missing values or duplicates, validating that the predefined range constraints are respected, check of value representation and data type constraints. The various dimensions that should be considered during the Data quality check can be grouped under the following categories: Accuracy (in what degree the data corresponds to the “real world” object correctly (e.g. football athlete’s age is over 50 years old), Completeness (whether the collected data are adequately complete in comparison to the potential of “100% complete” (e.g. critical data such as required fields of a questionnaire must have no missing values), Consistency: whether two or more representations of an object against a definition present differences (e.g. date format is MM/DD/YYYY in the USA and DD/MM/YYYY in Europe), Timeliness: whether the validity of data is related to a specific point in time (i.e. data can be valid for a specific time period and must be up-to-date when published), Uniqueness: whether information is only recorded once, based on specified unique identifiers (i.e. duplicate entities having different attribute values), Data Validity: whether the data conforms to the format, type and range as per its definition.

Data Transformation aims to ensure compliance with the available DataVaults schemas as well as the necessary data quality levels to achieve maximum usefulness and usability. In particular, Data Transformation incorporates: (a) data filtering, reducing the content of noise or errors and unmask important features from the data; (b) data cleaning to detect, modify and correct or remove dirty or coarse data, such as corrupt records, irrelevant or unnecessary parts of data, data not compliant to the defined constraints and formats etc.

2.2.2 Operation II.2: Semantic Enrichment & Annotation

Semantic Enrichment and Annotation is responsible for the attachment of additional, machine-readable information to the retrieved data, regarding various concepts (e.g. people, locations, organizations, etc.) and a multitude of domains (e.g. energy, health, tourism), as well as the extraction of the underlying semantics and the mapping of identified concepts to the DataVaults data schemas.

The information that is attached to the collected datasets, comes from vocabularies, ontologies and semantic models of the related domains. The mappings from the underlying models of various data sources to the common DataVaults schemas, have been defined by the DataVaults Data Scientist, taking into consideration all this domain-related information.

The overall process facilitates the unambiguous interpretation of data and maximises their reusability, as the modelling differences among the various data sources are bridged under the common, domain specific DataVaults models, while making use of domain expertise from external sources.

2.2.3 Operation II.3: Data Linking

Data Linking refers to the identification and linking of datasets from different data sources, about the same person or entity, to create a new, richer dataset or a collection of interlinked datasets in the same catalogue.

The task of determining whether two datasets can be linked to one another, to represent the fact that they refer to the same real-world Individual in a given domain or the fact that some kind of relation holds between them, is performed on the basis of the evaluation of the degree of similarity among different data instances. This task is called instance matching and includes techniques and (semi-)automated tools for performing the similarity evaluation task.

The data linking process includes the following tasks: 1) configuration – set up the parameters used during the instance matching step, in order to compare object description. 2) predicate selection - the choice of the predicates used to represent the meaning of the linking relation. 3) pre-processing & optimization - transform the original representation of data according to a reference format used for the comparison and minimize the number of comparisons that have to be executed in order to produce the final mapping set. 4) matching – the object description comparison according to the metrics chosen in the configuration step. 5) post-processing & validation - refine the matching result according to specific domain or application requirements and validate the mappings with respect to a formal definition of consistency for the mapping set.

2.3 PHASE III: ASSET STORAGE AT INDIVIDUAL'S SIDE (DATAVAULTS PERSONAL APP)

This phase pertains all operations for the secure storage of the personal data assets at the Individual's side. A data asset consists of the processed data that have been retrieved from a specific data source during Phase I, together with any metadata and semantic information that was linked to it during Phase II.

2.3.1 Operation III.1: Asset Generation and Storage

This operation concerns the execution of certain pre-defined data analytics algorithms and statistics at the users' side, with the aim to showcase to Individuals some illustrative parts of their data, and transform their data into assets, whether these are combined datasets, analysis outputs or visualisations. This step allows the user to grab a better understanding on some of his activity, as presented through intuitive and easy to understand graphs, and therefore acts as a motivation for the user to continue using DataVaults as he would be able to see his progress on certain domains of interest.

When it comes to the actual analytics which go a step further than the simple visualization and statistical analysis of the data shown to the user through the aforementioned graphs, it needs to be noted that due to the constrains that are imposed by the nature of the DataVaults Personal App and its underlying infrastructure, these are regarded as "edge analytics", implying the use of pre-defined algorithms, with little customization options offered to the user, and in principle consists of methods that are both performant and not resources hungry. The outputs of those operations can be saved as "assets", similarly to the data which might not be processed by the user through the edge analytics methods.

2.3.2 Operation III.2: Encryption at the Personal DataVaults Side

The secure storage of the data assets is facilitated by a cryptographic engine that will ensure both **information protection and secure data management** (covering the entire data lifecycle ranging from secure collection and storage, in the DataVaults Personal App, to secure sharing and storage in the DataVaults Cloud platform) and **privacy preservation** based on the user's preferences for future (anonymous and privacy-preserving) data trading transactions with interested Data Seekers. To this end, DataVaults will protect data and resources against leak or improper modifications, while at the same time will ensure data availability to legitimate users. Internal storage and ledger infrastructures, handling personal and/or corporate data, can track its provenance and will be regularly audited to comply with specified security and privacy policies and regulations.

For the former, the cryptography engine will be responsible for implementing all the core crypto primitives for protecting user's data from being compromised and tampered with. The integration of encryption technologies will also guarantee data access rights to only authenticated and authorized system users. Advanced hybrid encryption mechanisms will be integrated in DataVaults for protecting data confidentiality without infringing data access efficiency. For the secure storage, in the DataVaults Personal App, symmetric cryptography will be leveraged. This will revolve around the use of strong encryption ciphers (based on formally verified key derivation functions) during several operations such as authentication or transport sessions, and to also protect data that is stored on the user's device. The block cipher modes references in the current specification are defined in ISO/IEC 10116:2006. For the secure sharing and storage, in the DataVaults Cloud platform, Attribute-based Encryption (ABE) mechanisms will be provided. This provides encryption of data assets under a "data sharing policy", so that only Data Seekers, adhering to this policy, can be provided with the respective secret key for getting access to the target data resources. Besides data owner policy-based access control, ABE also ensures that user secret keys (used for the local storage)

will not be shared with third parties since the encrypted cloud storage will use unique attribute-based keys managed by the DataVaults ABE Trusted Component.

For the latter, depending on the defined data sharing configuration and selected user privacy level (Operation IV.1), privacy enhancement will be achieved (beyond the integration of traditional obfuscation schemes such as Digital Twins and User Personas[1]) through the use of trusted computing technologies (i.e., TPMs) as a central building block towards the provision of privacy-preserving signature schemes based on the use of Direct Anonymous Attestation [1]. DAA is a platform authentication mechanism that enables the provision of privacy-preserving and accountable services, thus, allowing for the *minimum disclosure* (of user personal information), *conditional (user) anonymity*, *unlinkability* and *forward & backward privacy*. It is based on group signatures which give strong anonymity guarantees. In particular, the use of DAA by DataVaults will achieve the following security and privacy properties:

- **User-Controlled Anonymity:** Identity of the users cannot be revealed when they anonymously share their data assets, with the DataVaults Cloud platform, for further data trading. Privacy-preserving credentials (i.e., pseudonyms) will be leveraged for the anonymous provision of data assets (if strong privacy level has been selected by the user). Data assets can then be seen to belong to pseudonyms (i.e., user pseudo-identities) so that appropriate compensation can be transferred to the users (linked to those pseudonyms) but no other identification will be feasible.
- **User-Controlled Linkability:** User controls whether their "sharing/uploading actions", of different data assets, can be linked together. If required, no different data bundles will be able to be linked to the user's id when shared with third-party Data Seekers.
- **Correctness:** All shared data assets will be able to be verified for enhanced data integrity and correctness.

2.3.3 Operation III.3: Local Asset Storage

Local Asset Storage operation encompasses the activities which lead to persisting the encrypted personal data assets at the Individual's side (i.e. the local storage). The retrieved personal data along with their metadata are stored, having already adhered to the pre-defined DataVaults schemas, enabling easy entry, querying and analysis.

2.3.4 Operation III.4: Local Asset Indexing

During the Local Asset Indexing operation, in order to facilitate better querying results and data retrieval efficiency, the DataVaults Secure Storage facility incorporates a data catalogue service for the indexing of the data and their metadata that have been stored to the local storage.

2.3.5 Operation III.5: Local Asset Deletion

The Local Asset Deletion operation concerns the process of completely removing a data asset which has been previously collected by DataVaults, upon the decision of the Individual. As this operation involves only the Individual's side and is independent of DataVaults Cloud Platform

[1] More information can be found in Deliverable D2.1 – "*Security, Privacy and GDPR Compliance for Personal Data Management*"

storage and any active contracts, it can be seen as a straightforward deletion procedure that is initiated by the Individual's request.

2.4 PHASE IV: ASSET SHARING TO DATAVAULTS CLOUD PLATFORM

This phase pertains to all aspects around uploading and sharing data assets through the DataVaults Cloud Platform. This includes various operations, from the configuration of sharing aspects that is performed by the individual to the actual sharing of their data to the DataVaults Cloud Platform utilizing contracts to ensure the trustworthy exchange of assets. Finally, a blockchain-based mechanism is foreseen for the effective implementation of the data providers' compensation for any of their data assets that are requested and acquired by Data Seekers.

2.4.1 Operation IV.1: Data Sharing Configuration

Whenever an Individual decides to make data available to Data Seekers through the platform, s-/he is prompted to define various details regarding data sharing. These span from selecting the anonymisation level (non-anonymised, anonymised as Digital Twin, etc.) to setting the price and choosing the licensing terms.

In particular, the sharing configuration consists, amongst others, of the following selection parameters:

a) Anonymisation Level Selection: In particular, the Individual can select whether they want to share their data eponymously or anonymously. In the second case, they are provided with two options: The first option is to share an anonymized version of their data, where all personally identifiable information has been altered to avoid identification; in other words, create a Digital Twin. The second option is to share their data only as part of anonymised user groups that are constructed by the DataVaults Data Scientist by aggregating multiple users' data, i.e. a Persona.
b) Visibility Level Selection: A second aspect regarding data sharing is related to the discoverability of shared data in search queries. This is closely related to the access policies selection, that will be described in the next point, as the Individual can define not only the attributes of users eligible for purchasing the full data asset, but also the users eligible to find the data assets in their search results and view the selected previews.
c) Access Policies Level Selection: The construction of the appropriate access policies will enable the Individual to define who should be able to access each of her/his data assets based on various attributes, such as the type of organisation the Data Seeker is related to, the type of personal data (e.g. health, social) and more. An indicative usage example could be that of a user who generally wants to share eponymously data but is reluctant with a specific type of organisations having access to her/his eponymous personal data. At the same time, this user is eager to provide them access to her/his Digital Twin. Such cases are facilitated through attribute-based access control mechanisms that enable the construction of fine-grained policies that will be bound to the data assets and will be resolved whenever an access request for the specific data asset is made. It should be noted that a user could upload data assets to

the DataVaults platform without allowing sharing with any Data Seekers in any form, by appropriately defining the access policies.

d) Pricing Selection: Afterwards, the Individuals are prompted to create the pricing scheme for each of the data assets they share. As the monetisation of data assets is closely related to the anonymisation level, the availability and demand for this kind of data, and other aspects, DataVaults will provide the Individuals with pricing suggestions that could help them in setting a viable price while maximizing their possible gains. The Individual sets the price tag for the shared data assets.
e) Licensing Selection: The Individual can select an applicable licencing scheme that will be applied whenever a Data Seeker purchases the specific data asset under the specific pricing. The Individual can define the license parameters, such as the expiry period, permitted purpose of use, sharing terms and more.

The user shall be offered with the option to automate the data sharing configuration procedure for data that are collected and updated in a recurring manner, through the option to define a sharing schedule for the specific data source.

Furthermore, the various parameters of a data asset sharing configuration can be modified in the future by the Individual. This involves changes in the access policies, the sharing schedule and more. A user can even completely revoke access to the shared data asset, by appropriately configuring the bound access policies. Of course, for the enforcement of these changes, DataVaults shall take into consideration any contracts in effect and ensure that Data Seekers that have purchased these data can acquire access for the respective period.

It should be noted that the Individual can share not only her/his personal data that have been collected from the connected data sources, but also any stored results from the Analytics and Visualisation Phase.

2.4.2 Operation IV.2: Sharing Risk Information

This operation is associated to raising the awareness of the Individuals on the privacy exposure impact of sharing data assets. The Individual is provided with a Privacy Metrics Dashboard that displays her/his current and projected risk estimations. These estimations have been calculated based on the data assets s/he has already shared, as well as on the data s-/he intends to share (if a sharing configuration is under design). More specifically, risk exposure metrics are calculated by taking into consideration all sharing aspects (anonymisation level, discoverability) as well as the information provided by the nature of the data itself and are updated whenever a modification to the shared assets occurs.

2.4.3 Operation IV.3: Private Contract Management – Individual & DataVaults Cloud Platform

This operation is responsible for all actions revolving around the creation and management of contracts between the Individual and the DataVaults platform. A contract is validated whenever a data asset or a cryptocurrency transfer takes place, in order to seal the exchange and guarantee the enforcement of the mutually agreed terms.

DataVaults offers the option for the sharing of data to the Cloud to be initiated either by the Individual on her/his own initiative, or by a Data Seeker upon request for data that are not yet

available through the DataVaults cloud. On the first case, the contract shall encompass the sharing configurations made by the Individual regarding the various sharing aspects (data asset visibility, pricing, licensing etc.). On the second case, the contract is constructed based on the offer made by the Data Seeker for the requested data asset and the issued contract shall be reviewed and accepted by the Individual for the data sharing to take place. Once the data asset is successfully stored on the cloud and the subsequent updates of the access policy and risk management engines are completed, the contract between the Individual and the Platform is validated and remains effective until the set expiration date is reached – if any applies.

Taking into consideration the right of the data owners to change their mind in the future regarding shared data assets, and at the same time respecting the indirect agreement that has been sealed through contracts between the Individuals and the Data Seekers, DataVaults has designed a data asset sharing revocation mechanism that ensures that both sides are covered. Thus, the contracts are considered more as a living entity. In case an Individual decides to stop sharing their data assets under the defined terms, or even to completely delete them from the cloud, DataVaults shall invalidate all private active contracts (e.g. between the Individual and the DataVaults Cloud Platform), and append new ones updated with the new terms, so that any new data asset transfers happen under the revised terms.

2.4.4 Operation IV.4: Compensation Management

A compensation mechanism will attach real value to purchased data assets. The implementation of a blockchain-enabled procedure is foreseen, providing an automated way for the completion of the financial transaction in the form of a cryptocurrency, without the need for further involvement of the two parties. Apart from the integrity and certification of the transaction, this mechanism ensures also the privacy and security of the Individuals and Seekers, as it is cryptographically secured.

A contract between the Individual and the DataVaults Cloud Platform shall be created upon successful completion of any data asset transaction between the DataVaults Cloud Platform and a Data Seeker. These transactions could be done without or with the interference of the Individual. Subsequently the Individual is compensated in the selected type of currency or any other, as defined by the pricing and sharing terms set in the data asset sharing contract.

2.5 PHASE V: DATAVAULTS CLOUD PLATFORM ASSET STORAGE

This phase pertains all operations which are applicable whenever a specific dataset is to be uploaded and stored from the Personal DataVaults App to the DataVaults Cloud Platform. Direct Anonymous Attestation is a cryptographic primitive that will be employed prior to any connection or data transfer, to enable remote authentication of the exchanging parties and ensure that the involved devices are not compromised.

2.5.1 Operation V.1: Encryption and Access Policies Enforcement

At this operation, the data assets of the Individual are encrypted based on the upload/sharing configurations of the Individual. In this context, Attribute-based Encryption (ABE) (Operation III.2) and Searchable Encryption (SE) mechanisms will be leveraged. As aforementioned, ABE

provides encryption of data assets under a “data sharing policy”, so that only Data Seekers, adhering to this policy, can be provided with the respective secret key for getting access to the target data resources. ABE guarantees the confidentiality of data but also provide data owner policy-based data access control so that the owner can decide who can access its data via specified sticky policies.

On the other hand, searchable encryption is one of the cryptographic schemes under consideration for the implementation of an Encrypted Searchable Data Lake, as it enables the execution of queries over encrypted data without revealing any plaintext information to the requesting user/application. In particular, ABE provides expressive and fine-grained data access control, while SE offers privacy-preserving data search (for the case where a data collector is an external system) and ensures dynamic encrypted data sharing. DataVaults will use lightweight but expressive policy-based ABE to ensure symmetric key be securely protected and be further efficiently shared under the corresponding policy. Additionally, SE is employed to guarantee that the external Data Seeker is not able to “see through” all the current data of the platform when doing data searching, and it will help this data collector securely decrypt – collect – the encrypted data it searches without endangering key secret exposure.

While this scheme is suitable for privacy-preserving information search and retrieval, it poses some performance and scalability challenges when it comes to the management of big data infrastructures. Towards this direction, the DataVaults Cloud platform will generate and manage a searchable encrypted index structure in combination with the distributed ledgers infrastructure (Blockchain-based mechanism). The platform will convert the most frequent metadata/search keyword/mode into the “secret information” which can be embedded into a 0/1 binary tree structure and store the location of the pointer on the ledger. By this secret twist, the index structure builds up a strong link with the ledger in such a way that the Data Seeker may only need to execute a privacy-preserving search over the structure to locate the specific data asset (leaf of the tree) and then it will obtain the location information of the encrypted pointer on the ledger. To do so, the Seeker must be given a search token by the DataVaults Cloud platform, which can be seen as an approval of the permission of searchability, helping the Seeker to find a correct path from the root to a specified leaf on the tree structure. The search token, however, will not allow the data Seeker to know anything except a location on the ledger. To synchronize the real-time ledger expansion, DataVaults will design a new type of dynamic SE mechanism that allows the system to build up encrypted index structure growing with the ledger. By using SE, DataVaults will not only provide privacy but also high efficiency (O(logN)) in search over massive amounts of data assets.

2.5.2 Operation V.2 DataVaults Cloud Platform Asset Storage

This operation pertains to the activities which lead to persisting the encrypted personal data in the form of data assets at the side of the DataVaults Cloud Platform. The retrieved personal data along with their metadata are stored, having already adhered to the pre-defined DataVaults schemas, enabling easy entry, querying and analysis.

2.5.3 Operation V.3: DataVaults Cloud Platform Asset Indexing

Asset Indexing operation facilitates better querying results and data retrieval efficiency over the DataVaults Cloud Platform database through the application of data indexing techniques over the data assets and their metadata. Data assets stored at the DataVaults Cloud Platform are associated with the appropriate indexes, enabling faster and more efficient queries. Different types of indexes can be used for this purpose, depending on the applicable data model and metadata.

2.5.4 Operation V.4: DataVaults Cloud Platform Asset Deletion

This operation concerns the process of the complete removal of a certain data asset that has been previously uploaded to the DataVaults Cloud Platform. In the simplest of cases, a certain data asset is not part of any active public contract, therefore it can be directly removed from the DataVaults Cloud Platform.

However, this operation might as well involve data assets which could belong to active public contracts between the platform and Data Seekers. In this case, prior to their deletion from the platform, DataVaults shall ensure the availability of this certain data asset to any eligible Data Seekers until the end of the agreed time period. For all other Seekers, these data assets will not be discoverable anymore through the search and exploration operations. An automated process should do some housekeeping after the expiration of each public contract, by deleting personal data assets which have been in the meantime (i.e. before contract expiry date) removed by the DataVaults Cloud Platform by the Individual who owns it.

2.6 PHASE VI: ASSET EXPLORATION & EXTRACTION

This phase is related to the efficient exploration of data assets that have been uploaded to the DataVaults Platform, through appropriate query mechanisms, while enabling Data Seekers to actually acquire explored data of their interest through data provision services that enforce any access policies bound to the data by the provider.

It is strongly connected to Phase V, as it is responsible for the searching and extraction of data assets that have been persisted in the DataVaults data storage at this preceding phase.

2.6.1 Operation VI.1: Asset Searching

The searching and retrieval of data assets from the DataVaults storage takes place in this operation. The Data Seekers are able to search over the encrypted and the unencrypted DataVaults storage lakes for assets shared by Individuals, and retrieve any related information they are eligible for according to the access policies in place, such as a preview of the data asset, any descriptive metadata and its sharing properties (price, period, access policies). Various types of queries are supported by the search mechanism to maximise the discoverability of the data. Name search can vary from targeted searches for specific assets using exact name search, to keyword search for the exploration of the database. The stakeholder can also search for data assets based on metadata (ex. Location, age span, etc.), sharing settings (e.g. eponymous or anonymised) and pricing. The search mechanism of DataVaults supports also range queries to retrieve data that lie between the set boundaries.

The Personas that have been created by DataVaults Data Scientists are also available through the asset search feature.

2.6.2 Operation VI.2: Public Contract Management - Data Seeker – DataVaults Platform

Whenever a Data Seeker decides to acquire a data asset that is available through the DataVaults cloud, a contract is issued between the two parties. This contract entails the terms defined by the data asset sharing configuration and is validated after the data asset has been transferred to the Data Seeker and the related cryptocurrency transaction has been completed successfully.

2.6.3 Operation VI.3: Asset Export

This operation allows the Data Seeker to export and download data assets acquired over the DataVaults Cloud Platform. Data Seekers can download either data assets they own, such as data they have uploaded to the platform and the outcomes of analytics in the Analytics sandbox, or data assets they have acquired from Individuals and they are eligible for, as defined by the related contracts and access policies in effect. This service is enabled by the available DataVaults APIs, which are designed with flexibility in mind, providing the stakeholders with the option to download the data assets in the format of their choice.

2.7 PHASE VII: DATA ANALYTICS

The Data Analytics phase is responsible for providing the appropriate tools to the DataVaults stakeholders to maximise the value of their data, by utilizing popular algorithms and techniques. The offered analytics span from basic statistics to knowledge extraction, enabling stakeholders to gain useful insights from their data. This phase entails also the visualization of data in meaningful graphical representations, like charts, graphs and maps, to provide an intuitive way to see and understand trends, outliers and patterns in data. Stakeholders are able to combine and link datasets from different sources during these operations to explore correlations and perform various tasks, like classification and regression. The analytics and the visualization operations are bilaterally connected, as it may occur that the Visualization results shall be fed back to the Data Analysis operation for further processing and vice versa. The results of the analysis tasks and the selection of the algorithms in the algorithm chain, along with any parameterization are saved for future use and can also be exported by the user. One aspect that shall be taken into consideration, are the computational limitations that may apply. These limitations can result in a differentiation between the analytics features that are available in terms of processing power.

2.7.1 Operation VII.1: Cloud Based Data Analysis

In the Data Analysis operation, the available data analytics algorithms can be applied - combined or on their own - on selected data assets, to create new knowledge and enable the user to perform processes such as clustering, classification, outlier detection and more.

For this, the project will make use of the Toreador framework [3] developed by ATOS. The Toreador framework is a set of tools that will allow a multitude of well-known methods to be offered to the DataVaults stakeholders, including among others: a) machine learning algorithms such as regressions, support vector machines, and k-means clustering; (b) deep

learning capabilities capitalising on the BIgDL framework [4], wherein deep learning libraries will be integrated with Spark; and (c) prescriptive analytics methods, to allow the extraction of patterns and the execution of "what-if" simulation scenarios based on the existing data that is available. A summary of the algorithms that will be provided can be found in the table below.

Table 1 - Toreador Algorithms

	MLlib[2]	H_2O[3]	FlinkML[4]
Interface language	Java, Python, Scala	Java, Python, R, Scala	Java, Python, Scala
Associated platform	**Spark, H2O**	**H2O, Spark (Sparkling Water[5]), MapReduce**	**Flink**
Current version (23/07/2020)	**3.0.0**	**3.30.0.7**	**1.4**
Graphical user interface	–	✓	–
Classification and regression algorithms			
Decision tree	✓	–	–
Logistic regression	✓	✓	✓
Naïve Bayes	✓	✓	–
Support vector machine	✓	–	✓
Gradient boosted trees	✓	✓	–
Random forest	✓	✓	–
Generalized linear model	–	✓	–
Linear regression	✓	✓	✓
Stacked Ensembles	–	✓	–
XGBoost	–	✓	–
Clustering algorithms			
k-Means	✓	✓	✓
Fuzzy k-means	–	–	–
Streaming k-means	✓	–	–
bisecting k-means	✓	–	–
Power iteration	✓	–	–
Latent Dirichlet allocation (LDA)	✓	–	–
Collaborative filtering (cf) algorithms			
Alternating least squares	✓	–	✓
Dimensionality reduction and feature selection tools			
Principal component analysis	✓	✓	–
Singular value decomposition	✓	–	–
Chi squared	✓	–	–
TF-IDF	✓	–	–
Word2Vec	✓	–	–

[2] http://spark.apache.org/docs/latest/mllib-guide.html

[3] http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science.html

[4] https://ci.apache.org/projects/flink/flink-docs-release-1.3/

[5] http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html#sparkling-water-users

Polynomial Features	–	–	✓
Standard Scaler	–	–	✓
MinMax Scaler	–	–	✓
Additional algorithms			
Association rule learning	✓	–	–
Deep learning	–	✓	–
Topic modelling	✓	–	–
CEP (Complex Event Processing)	–	–	✓

Given the nature of the Toreador framework and the possibility to deploy the algorithms where needed by just changing some configuration variables, a sandboxed environment, appropriate for experimentations, is foreseen. There, the user shall be provided with the option to discard the analysis' results and completely erase any traces of data once s-/he has completed the analysis, in case s-/he does not wish to save the results or the configuration details for future use.

Once the DataVaults stakeholder selects the algorithm to be utilised over a specific dataset, the analytics engine shall assess their compatibility in terms of parameters such as data type and format and ensure that the selected operation is feasible.

It is possible that for the proper operation of the algorithms, the data must be pre-processed before being fed to the analytics engine, in order to come in the required format, type, range. For this purpose, the Toreador framework allows the integration of personalized python algorithms to be run on top of the available platforms in the framework, hence some pre-processing algorithms will be provided to transform the datasets from the DataVaults data model into a dataset that can be processed by the algorithm selected by the user. In case of a permitted operation, the user is prompted to configure all relevant analysis parameters (e.g. number of clusters, accuracy etc.) before the computation is performed. This is allowed by using a combination of Ansible [6]and Docker[7]. A predefined set of Ansible templates is provided by Toreador, containing the most used algorithms that can be combined to create a playbook. When executed, it creates a set of Docker images to deploy the platforms required by the algorithms in the first place (Hadoop [5], Spark [6], etc.), and then deploys the algorithm to be run on that previously deployed platform.

The analytics playground allows also for the successive execution of algorithms over a dataset, through the definition of algorithm chains. In this case, the result of the execution of one algorithm is handed over to the next algorithm in the row and so on. However, as it is possible for the accumulative effect of many possible sequences of algorithms to not make any sense, DataVaults provides pre-defined suggestions for classic combinations of algorithms that work well together.

[6] https://www.ansible.com/

[7] https://www.docker.com/

2.7.2 Operation VII.2: Data Visualisation

The graphical representation of data through Data Visualisation is a key enabler in understanding their nature and any underlying trends, especially in the case of big data and of non-easily comprehensible data. Various visualisation techniques are offered to the DataVaults stakeholders to choose from, based on their needs and the nature of the data at hand. The selection of the most fitting technique for the intended application is a challenge and may require the experimentation with various visualisations, until the best option is found. An intuitive dashboard shall be provided for the customization of visualisation parameters, as well as for the direct interaction with the graphs and plots to explore them in various levels of granularity, apply filters, add or remove layers and more. The visualisation results constitute data assets that can be saved for future use, be integrated in interactive reports and be exported to be used locally.

2.7.3 Operation VII.3: Persona Creation

Another functionality available to DataVaults Data Scientists, is the creation of personas. The Scientists are able to use the Anonymisation Bundle, to group data coming from different Individuals and process them using statistical methods, in order to create 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.

These Personas can provide valuable information to Data Seekers while preserving the anonymity of the indistinct Individuals that have been considered for the specific modelling. The Personas created by the DataVaults Data Scientists are also stored on the DataVaults Cloud Platform and are available to Data Seekers under a price set by the DataVaults Data Scientist.

2.8 PHASE VIII: ADDED VALUE SERVICES

2.8.1 Operation VIII.1: Notifications

DataVaults aims to provide notification services to the Individuals and Data Seekers taking part in the exchange of personal data assets, as it is useful to inform the users of the status of an operation or to trigger them into performing further actions. Thus, notification services will be provided at all DataVaults Phases facilitating the interaction of the users with DataVaults. More precisely, this service may allow any Data Seeker to post requests for specific assets and the corresponding Individual may be notified through the DataVaults Personal App. Similarly, Individuals may receive notifications from the DataVaults Cloud Platform regarding their current privacy risk exposure. Both Individuals and Data Seekers may be notified about the success of transaction operations regarding contracts. As a rule of thumb, automated communication via notifications needs to be carefully monitored in order not to overwhelm the asset providers and consumers with irrelevant notifications.

2.8.2 Operation VIII.2: Sharing Gains

The Individual can view at any time the actual value s-/he has gained from her shared data assets that have been purchased by Data Seekers, in the form of cryptocurrency.

3 DATAVAULTS HIGH LEVEL PLATFORM OPERATION SCENARIOS

This section aims to present the core operations that a platform following the above mentioned methodology could execute. Initially, the presentation of the five main Actors who are taking part in the ecosystem of DataVaults is provided.

Following the identification of the main actors, eleven high-level usage scenarios of DataVaults are presented. All of the high-level usage scenarios are based on the DataVaults methodology, DataVaults demonstrators and key findings from D1.1 "DataVaults Data Value Chain Definition". The high-level scenarios were initially drafted as workflow diagrams (i.e. clear sequences of steps) that one or more Actors would follow in order to achieve certain objectives. After collecting feedback on the initial drafts, the outlines of eleven scenarios were chosen as representative of all different prospected users that correspond to envisaged DataVaults utilization cases relevant to the stakeholders' needs.

3.1 ACTORS DEFINITION

In this Sections, the five main Actors of DataVaults are identified. Apart from the Individual and the Data Seeker who represent human Actors, we have described the DataVaults Personal App and the DataVaults Cloud Platform as separate Actors, as in some cases the two must work together to achieve the desirable result. Finally, the DataVaults Cloud Scientist has been defined as a human Actor who is familiar with big data analytics and aware of algorithms and statistics for data analysis and can therefore tweak the platform as necessary.

3.1.1 Individual

An Individual is a private person who generates and collects their own personal data from various services, devices and applications. An Individual uses the DataVaults Personal App to construct their unified personal data hub and collect at a single place all their personal data in a secure and trusted manner. At the same time, an Individual manages their personal data and decides what to share and with whom, receiving compensation for the data assets they place at the disposal of third parties. An Individual connects to the DataVaults Personal App which is installed on a given device, or directly via web-based access and its main activities include collecting personal data, configuring sharing parameters for those data, as well as analysing them at a basic level.

3.1.2 Data Seeker

A Data Seeker (also titled as 1st-tier Economic Operator, as per the DoA) is an organisation of any type, public or private. A Data Seeker is interested in acquiring Primary Personal Data from Individuals and create business intelligence based on it. A Data Seeker has or wishes to have the ability to combine Primary Personal Data (e.g. social media activity) with other types of data they already possess (e.g. demographic data) with a goal to create new datasets or relevant derivatives (insights, reports, etc.). A Data Seeker connects to the DataVaults Cloud Platform via an API or via web-based access and its main activities include exploring, acquiring and analysing Primary Personal Data from Individuals.

3.1.3 DataVaults Personal App

The DataVaults Personal App or Personal DataVault (as per the DoA,) is the personal side of DataVaults, which resides at the side of Individual users. The DataVaults Personal App retrieves personal data from various data sources (services, devices and applications), transforms them and stores them locally. The DataVaults Personal App offers capabilities to the Individual user to manage data access policies, configure data sharing parameters, analyse and visualize their data and remain aware of privacy exposure. Finally, the DataVaults Personal App connects to the Private Ledger of the Individual user and interacts with the DataVaults Cloud Platform.

3.1.4 DataVaults Cloud Platform

The DataVaults Cloud Platform is the central part of the DataVaults architecture, residing in the cloud. The DataVaults Cloud Platform stores personal data of Individual users in the cloud upon user selection, indexes them and creates an Encrypted Searchable Data Lake, which includes metadata and data samples and allows the operation of certain searchable encryption. The DataVaults Cloud Platform anonymises personal data of Individual users by implementing the Digital Twin Generator and the Persona Group Generator. The DataVaults Cloud Platform allows Data Seekers to search through encrypted and anonymised personal data of Individuals and express their interest to acquire them. At this point, the DataVaults Cloud Platform composes and validates contracts, in order to grant access to the data assets for the Data Seekers and to compensate the Individual who provided them. Finally, the DataVaults Cloud Platform allows Data Seekers to explore and analyse data assets and experiment within the DataVaults platform, by combining the extracts of personal data with their own (private) data and by running various analytics.

3.1.5 DataVaults Data Scientist

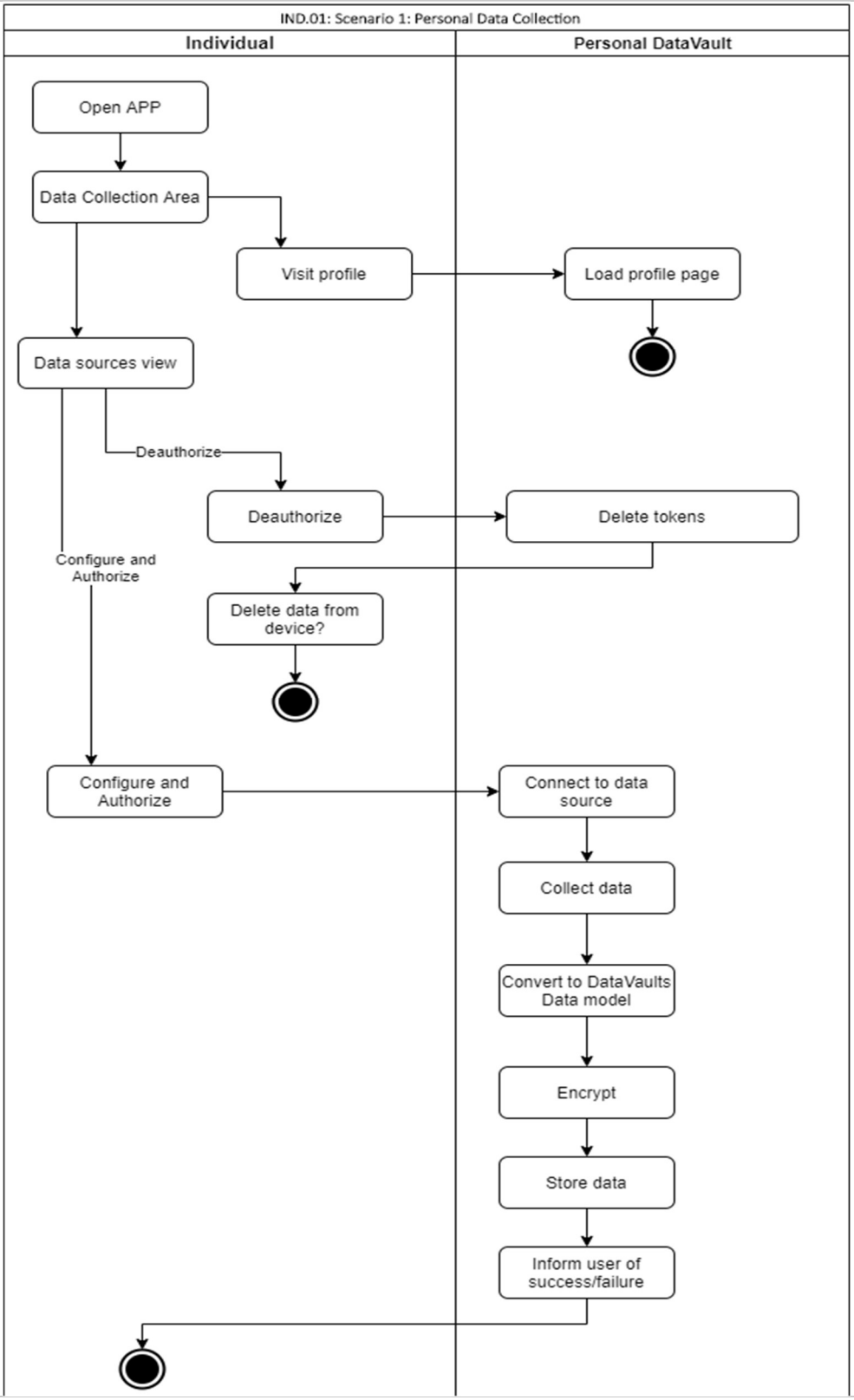
A DataVaults Data Scientist is a technical user who is familiar with big data analytics and aware of algorithms and statistics for data analysis. A DataVaults Data Scientist is well trained on the use of analytics engines.

3.2 SCENARIOS DRIVEN BY INDIVIDUALS

Six operation scenarios that evolve around the Individuals, are detailed in this Section. The scenarios present the steps that are taken by the Individual when handling their personal data, from the configuration of the collection details, the exploration and analysis of data in the Personal DataVaults App, as well as the various aspects of sharing their data to the DataVaults Cloud App.

3.2.1 Scenario 1: Personal Data Collection

Scenario ID	IND.01
Scenario Title	Personal Data Collection
Scenario Actors	Individual, DataVaults Personal App
Overview	An Individual uses the DataVaults Personal App: ● to manually update his/her profile data, and

	• to collect his/her personal data from a specific data source, following a process of authorising the App to perform the necessary data fetching. Afterwards, the data is collected, is matched to the DataVaults data schemas and is stored securely at the Individual's side.
Triggers	An Individual uses the DataVaults Personal App with the intention either to update his/her profile data or to authorise/deauthorise the DataVaults Personal App to collect his/her personal data and store it locally and securely.
Workflow	

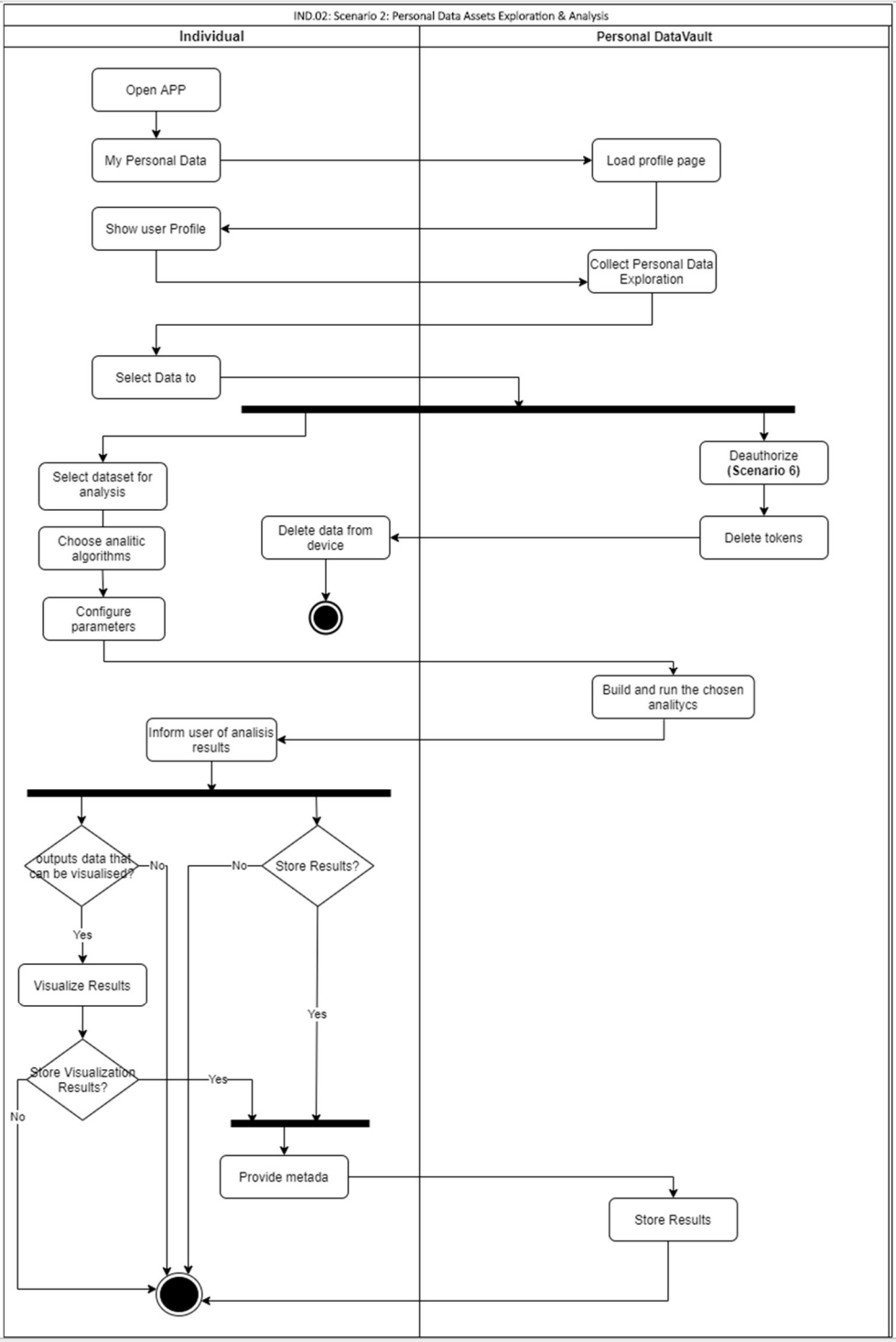
Scenario Sequence	1. The Individual opens/logs into DataVaults Personal App and visits the data sources management area. 2. The Individual views by default, upon visiting the data sources management area, all data sources (already connected to DataVaults Personal App or not) available on this device as well as his/her profile data. 3. The Individual chooses to: a. [*alt-flow 1*] Visit his/her profile area where s/he can update/add new data that are relevant to his/her profile (demographics, etc.), likings, personal preferences, etc. **The Usage Scenario ends here.** b. [*alt-flow 2*] Select one of the data sources provided to him/her by the DataVaults Personal App (i.e. wearable API, smart home device API, other smartphone app already on the device, etc.) and chooses to: i. [*alt-flow 2.1*] authorise the DataVaults Personal App to collect data from this data source by either a) creating a connection between the DataVaults Personal App and the data source API/webservice by providing his/her credentials of that service), or b) asking the DataVaults App to seek authorisation from an App to retrieve data that is already installed on his/her device and ii. configure some options which are relevant to this authorisation process and data retention (e.g. token expiration date, interval of data collection, frequency of data collections, auto-deletion of obsolete/expired data, auto-renewal of the operation upon token/data asset expiry, etc.)., OR iii. [*alt-flow 2.2*] deauthorize the DataVaults Personal App from collecting data from this data source. 4. [*Continuing from alt-flow 2.1, operation 3.b i&ii*] The DataVaults Personal App initiates the process to connect to those data sources based on the configuration compiled in the previous operation. 5. The DataVaults Personal App collects data from the selected data source (Data Fetcher) and prepares the data based on pre-predefined data management operations for the selected data source (data cleaning, semantic enrichment, etc.). 6. The DataVaults Personal App matches the collected data to the existing DataVaults Data Schema (Data Transformation). 7. The DataVaults Personal App encrypts collected data (Encryption/Decryption Engine). 8. The DataVaults Personal App securely stores data at the Individual's side. 9. The Individual is notified about the success of this operation and the new data source is displayed in his/her dashboard as "connected", alongside with some statistics. **The Usage Scenario ends here.** 10. [*Continuing from alt-flow 2.2, operation 3.b iii*] The DataVaults Personal App initiates the process of de-registering/erasing the tokens it already has stored for the specific data source.

	11. The Individual may choose whether to delete the data from that data source already stored on his/her device or not (jump to operation #4 of Scenario 2: Personal Data Assets Exploration & Analysis).
User Benefits	The Individual is able: ● to view all data sources (already connected to DataVaults Personal App or not) available on his/her device. ● select which data sources s/he wants to connect to DataVaults and which data sources s/he wants to disconnect from DataVaults. ● retrieve data from external sources and collect them in one place.
Challenges	● Connections of the DataVaults Personal App with several different API's which are not always homogenised. ● Terms of use for third party APIs are changing constantly and may not allow data retrieval. ● Frequency of data fetching might be a problem due to free API service limitations. ● Storage capacity at user's side might be insufficient for the data volumes to be fetched. ● Data Retrieval needs in many cases applications to operate in the foreground.
GDPR Issues	● Especially in Scenario Sequence 3 (SS 3), full transparency on personal data collection and processing must be ensured for the lawfulness of the operations. It is necessary that the Privacy Notice provides the minimum list of information laid down in Art. 13 GDPR: the form and extent of information, as for instance in terms of language and granularity, must be adequate to the information's recipient. Transparency represents a prerequisite for the Individuals to exercise their rights under Chapter 3 of GDPR, including control and intervention. ● The consent should be managed in a fine-grained way, allowing distinct consent options for distinct processing operations/datasets/data sources, as well as partial granting or withdrawal of consent in some circumstances (SS 3b). Ongoing consent should be ensured and information on the privacy risk exposure should be adequately provided to the Individuals, besides statistics (SS 9). It is important both that Individuals can easily withdraw consent with undue effort and without detriment (through modalities as simple as those in place for providing consent) and that this choice is effectively and timely enforced (SS 3 b iii and 10, 11). Considering the wide range of data sources, a filter on the special categories of data listed in Art. 9 GDPR should be conceived, thus distinguishing between consent requests on "normal" personal data and those involving sensitive data. ● The User Interface should be user and data protection friendly, capable of facilitating as much as possible user control over his/her personal information and the granularity of informed consent (SS 3 b). ● In line with the accountability principle, the documentation and demonstration of compliance with GDPR throughout the whole process is important (workflow, SS 1-11).
Ethical Issues	● Need to seek for a trade-off on consent modalities to ensure that the consent is genuine, effective, and freely given, avoiding the so-called "consent-fatigue", as well as the inappropriateness of globalized and generic consent for multiple vague purposes. The future purposes of

	secondary use of the data by the Data Seekers might not be known at the time of the data sharing. ● To maximize user control and the "sharing the wealth" paradigm aimed at letting Individuals benefit from the value created by the use of their personal data in a more balanced and transparent way, data portability-driven tools should be foreseen for giving to individuals the access to/export their own data in portable, interoperable and machine-readable (usable and reusable) format, as well as the ability to further process their own data and take advantage of additional services (other than DataVaults' ones) for analysing them and drawing useful conclusions.
Privacy Issues	• Data Privacy: The Data Transformation, to be performed by the DataVaults Personal App (SS 3), should prepare the data in a privacy-preserving manner depending on the user's selected privacy requirements so as no user-sensitive information can be leaked when data is then transmitted to the DataVaults framework. This will be achieved through the Data Anonymizer agent running on the user's device employing any of the DataVaults privacy operation bundle suite including DAA, transformation to data personas, etc. (as described in D2.1).
Security Issues	• Configuration integrity and operational correctness of the DataVaults Personal App, running on the user's device, so as to provide verifiable evidence that the app collecting the user personal data (from the data sources) is not compromised, thus, leading to possible leakage and manipulation of the extracted information (SS 3 - 9). This will be achieved through the Attestation Agent (and the provided remote attestation algorithms; i.e, DAA) leveraging the trusted component hosted in the user's device. • Accountability and non-repudiation of the user's actions regarding providing the necessary privileges to the DataVaults Personal App to collect the user data from the selected data sources based on their preferred data sharing schema (SS 3 - 4). This will be achieved through the provision of appropriate signature mechanisms for verifying the correctness of user's actions. ● Confidentiality of the collected data to be stored locally at the user side (SS 8). This will be achieved through the use of strong crypto primitives (i.e., AES) by the DataVaults Personal App. Key management of the respective secret keys will be achieved through the host trusted component.

3.2.2 Scenario 2: Personal Data Assets Exploration & Analysis

Scenario ID	IND.02
Scenario Title	Personal Data Assets Exploration & Analysis
Scenario Actors	Individual, DataVaults Personal App
Overview	An Individual can browse through his/her own data and perform experiments with his/her own data assets; creating simple analyses, visualizing them and getting a holistic picture of which data assets the Individual has shared over DataVaults.

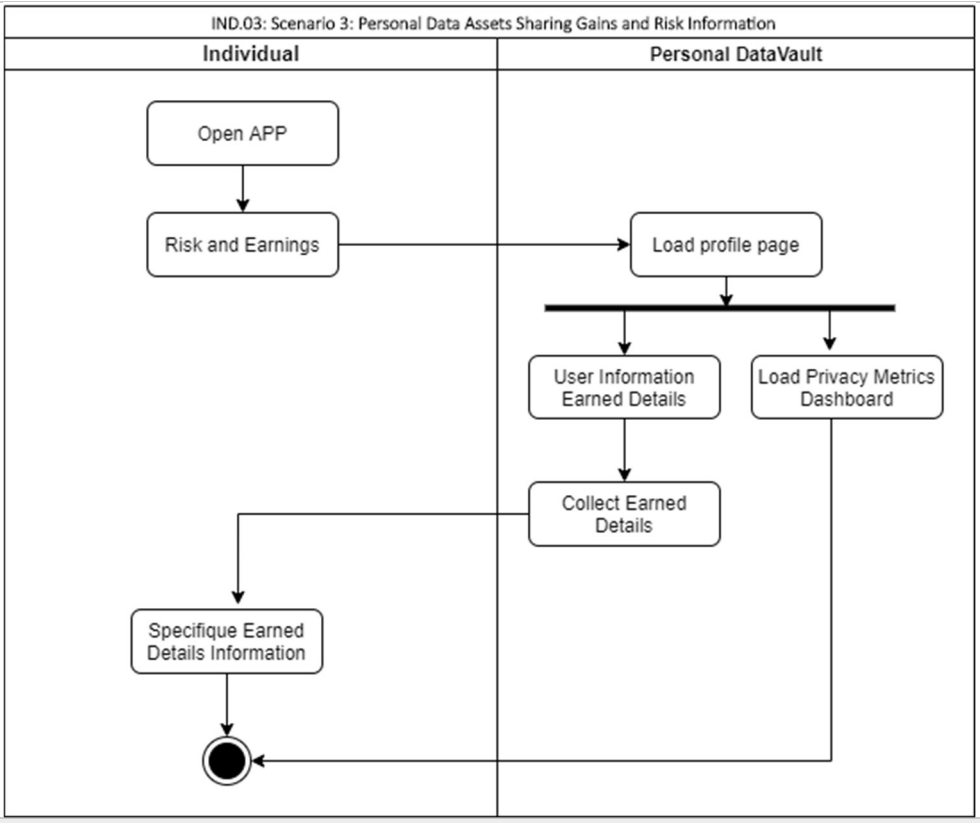
Triggers	• An Individual uses DataVaults Personal App to browse through, analyse and visualise their own personal data. • A request from a Data Seeker arrives asking for specific assets to be shared (see Scenario 9: Acquire Data Assets from a DataVaults Individual User).
Workflow	
Scenario Sequence	1. The Individual opens/logs into DataVaults Personal App and visits the personal data area. 2. The Individual can view his/her profile data that s/he has provided manually.

	3. The Individual can view all personal data collected from data sources authorized to work with DataVaults Personal App (i.e. all data assets), alongside with some statistics (e.g. volume, last retrieval date, etc.). 4. The Individual can select data which s/he would like to remove from his/her DataVaults Personal App: a. The DataVaults Personal App initiates the process of de-registering/erasing the tokens it already has stored for the specific data source, thus deauthorizing itself from fetching these data in the future. b. The Individual may choose whether or not to delete the data from that data source already stored at the Individual's side. 5. The Individual can select some data to apply simple analysis. This is done by: a. Selecting one or more datasets of the displayed data. b. Choosing the analysis algorithm to be applied on these data. c. Configuring some extra parameters (in case these are available). 6. The Individual is notified about the success of the data analysis job and is presented with a new dataset. 7. The Individual can opt to store the analysis in his/her DataVaults Personal App by providing some metadata (e.g. title, description, etc.). 8. The Individual can opt to visualise the analysis' results (in case the algorithm outputs data that can be visualised) and even store the visualization.
User Benefits	● The Individual is offered visualization and analysis of his/her personal data connected with DataVaults. ● The Individual remains aware and informed about his/her personal data and profile data.
Challenges	● Perform analysis and visualisation on a portable device could have a cost on the battery life. ● Advanced data analytics require specific libraries to be used in portable devices. ● Analysis notebooks are not intuitive for end users and are hard to use by non IT-literal users.
GDPR Issues	● Many remarks, such as those on transparency and privacy notice, are common to Scenario 1. ● The User Interface should be user-friendly and facilitate the functionalities foreseen by this Scenario, including the use by non IT-literal users, limiting discriminatory effects due to this kind of possible Digital Divide.
Ethical Issues	● This scenario emphasizes the "sharing the wealth" paradigm. In line with the ethics-related remarks on Scenario 1 regarding such paradigm and the user control, it is advisable to foresee tools for implementing at the maximum extent the data portability right, giving to Individuals also the chance to export their own data in portable, interoperable and machine-readable (usable and reusable) format, so that they can further process their own data also through other services/applications other than DataVaults' ones for generating high-value insights. In other words, this would widely allow Individuals to use the data for their own purposes, moving ahead towards the win-win data sharing ecosystem envisaged by BDVA for unlocking the social value of personal data, whilst boosting the business and economic value of personal data.

Privacy Issues	• Data Privacy: Users should be able to verify that their data, stored on DataVaults, have been collected following the security and privacy policies that were selected by the user when granting authorization to the DataVaults Personal Ap (SS 3). This will be achieved through the privacy risk assessment, that will be performed by the DataVaults Risk Assessment framework (during run-time), calculating a "privacy risk" quantification value, based on the user's stored data; this value should be at an acceptable threshold as was defined in the privacy policies. • User anonymity: User actions regarding any requested analysis (on the stored data) should not breach the anonymity level that was defined in the user privacy policy (SS 5). This entails that no additional personal information should be outputted by the analysis performed by DataVaults that can be used by an Observer to infer private information and whether the user performed a specific action. This will be achieved through the employment of appropriate differential privacy algorithms (i.e., k-anonymity) on the data yielded from the analysis tasks.
Security Issues	• Confidentiality and Authentication: Only the correct user, as the owner of the personal data, should be able to have access to this DataVaults personal area and perform data analysis tasks. This will be achieved through the appropriate user authentication and access control mechanisms, employed by the DataVaults Personal App, when logging in to the framework. Certificates will be leveraged through the host trusted component. • Non repudiation: Users should not be able to dispute a data analysis action that was initiated by them that requires some form of data processing. This will be achieved through the use of appropriate signature mechanisms.

3.2.3 Scenario 3: Personal Data Assets Sharing Gains and Risk Information

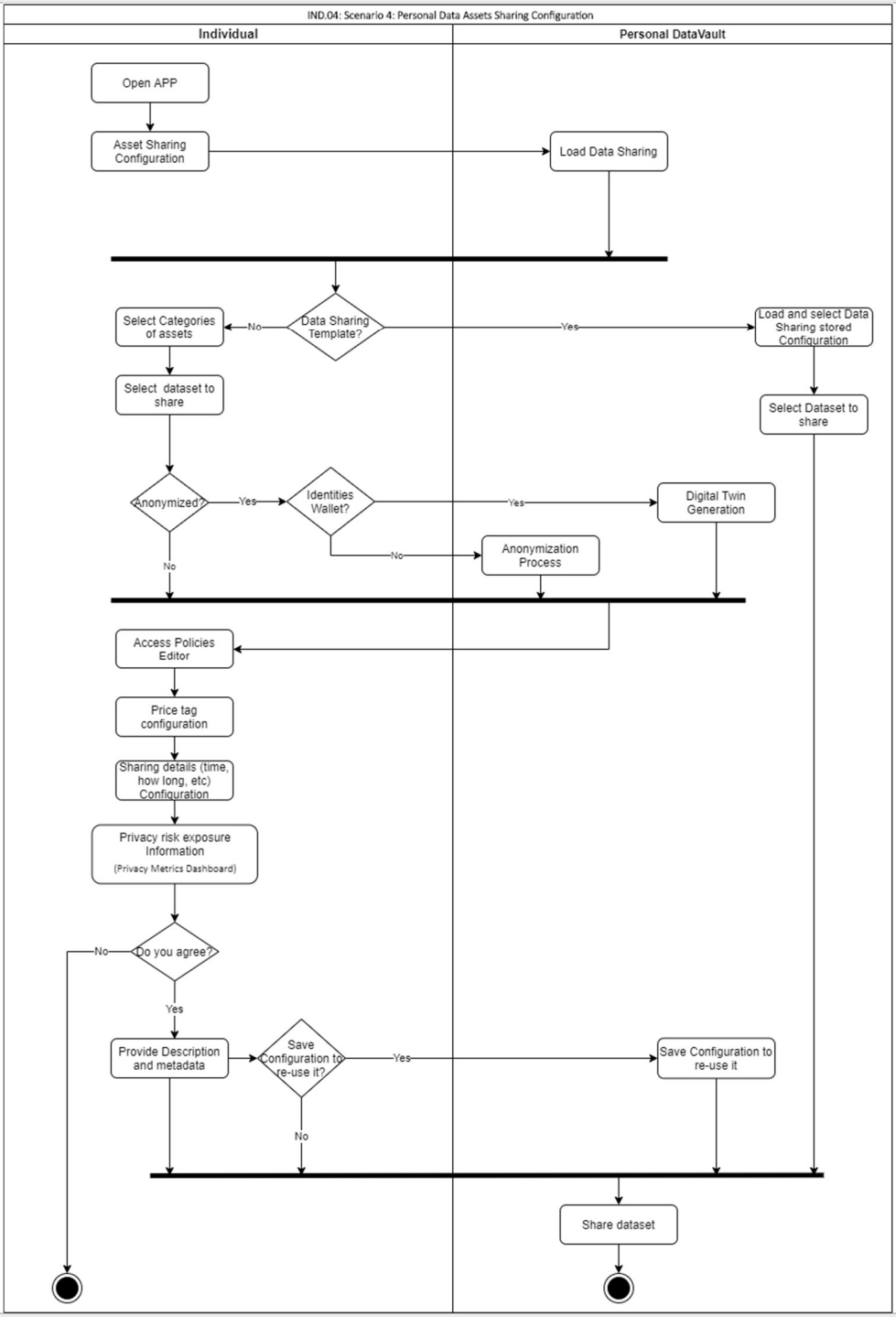
Scenario ID	IND.03
Scenario Title	Personal Data Assets Sharing Gains and Risk Information
Scenario Actors	Individual, DataVaults Personal App
Overview	An Individual is informed about what s/he has earned from sharing his/her personal data with DataVaults platform and what his/her current privacy risk exposure is.
Triggers	● An Individual uses DataVaults Personal App to learn about the transactions he already completed over DataVaults and current privacy risk exposure. ● A transaction is verified and a notification for updated risk exposure metrics arrives.

Workflow	
Scenario Sequence	1. The Individual opens/logs into DataVaults Personal App and visits the personal assets information view. 2. The Individual is informed about his/her current privacy risk exposure (Privacy Metrics Dashboard) based on the assets s/he has already shared over the DataVaults Cloud Platform. 3. The Individual is informed about what s/he has earned through DataVaults (e.g. amount of (crypto)currency earned) for each transaction. 4. The Individual can view the specifics of each transaction that has been already verified by the system.
User Benefits	● The Individual is offered information regarding his/her risk exposure. ● The Individual is offered information about his/her Personal DataVaults Wallet which contains his/her earnings based on the assets shared. ● The Individual is presented with a list of transactions verified.
Challenges	● Risk-Exposure Information analysis is difficult to be performed real-time.
GDPR Issues	● Considering the transparency requirement as a prerequisite for Individual's control and exercise of his/her rights in general, the timely provision of information on risk exposure is paramount. Therefore, in order to improve the effectiveness and usefulness of the Privacy Metrics Dashboard for transparency, Predictive analytics techniques/machine learning or other techniques should be applied also before the effective sharing of the assets over the DataVaults Cloud Platform (SS 2).
Ethical Issues	● Digital Divide (as mentioned in Scenario 2) and the possible resulting discrimination between Individuals as regards sharing attitude should be considered. ● Personal DataVaults Wallet should be suitable for different kind of rewarding incentives and compatible also with non-monetary compensation schemes and related earning (SS 3).

	• In line with the Fairness by Design principle, on which DataVaults is committed to rely, mitigation measures and safeguards should be taken also to foster societal fairness and equal opportunities, avoiding cases where people are deceived or unjustifiably impaired in their freedom of choice. This includes for instance to prevent that the poorest brackets of the population are disproportionately motivated to share own personal data.
Privacy Issues	• User Forward Privacy: The verification of the presented transactions (by the user – see Transaction Verification security property) should not breach the user defined privacy requirements (SS 4); i.e., it should not affect user unlinkability (if selected) by enabling an Observer to link this transaction directly back to the user. This will be achieved by the DataVaults ledger security mechanisms and the user's Personal DataVaults Wallet that can provide user anonymous authentication (i.e., DAA) when accessing the DataVaults ledgers. • User Privacy Exposure: The current user privacy risk exposure, as calculated by the DataVaults Risk Assessment framework, should not breach the privacy requirements as defined by the user. This requires that quantified privacy risk exposure values are kept within the user acceptable boundaries; in any other case, DataVaults should inform the user of appropriate actions that should be taken for privacy enhancement. This also pertains to no malicious user or Observer gaining a significant advantage of identifying (de-anonymizing) the user by exploiting the trading mechanisms; linking number of transactions with (e.g.,) amount of (crypto) currency earned by an individual
Security Issues	• Transaction Verification: Users should be able to verify the correctness of all (related) transactions presented by the DataVaults framework so as to have strong guarantees that these were executed correctly based on the agreed policies enforced through the smart contracts. This pertains to both the level of data sharing, included in the transaction, as well as the correct user compensation. It will be achieved through the DataVaults distributed ledgers and the provided management and access interfaces; i.e., the ID (or link) of the transaction should be also presented to the user so as s/he can check and verify the respective transaction block stored in the DataVaults private ledger. • Ledger Security: All transactions that have been verified by the system should not be tampered with when stored in the DataVaults ledgers. This will be achieved through the DataVaults Blockchain verification and validation mechanisms of all recorded transaction blocks.

3.2.4 Scenario 4: Personal Data Assets Sharing Configuration

Scenario ID	IND.04
Scenario Title	Personal Data Assets Sharing Configuration
Scenario Actors	Individual, DataVaults Personal App
Overview	This scenario describes all actions that an Individual performs to create a configuration based on which their data assets will be shared on the DataVaults Cloud Platform, ranging from data asset type selection, the privacy settings selection and the anonymisation mechanisms used, access policies design, asset value description, etc.

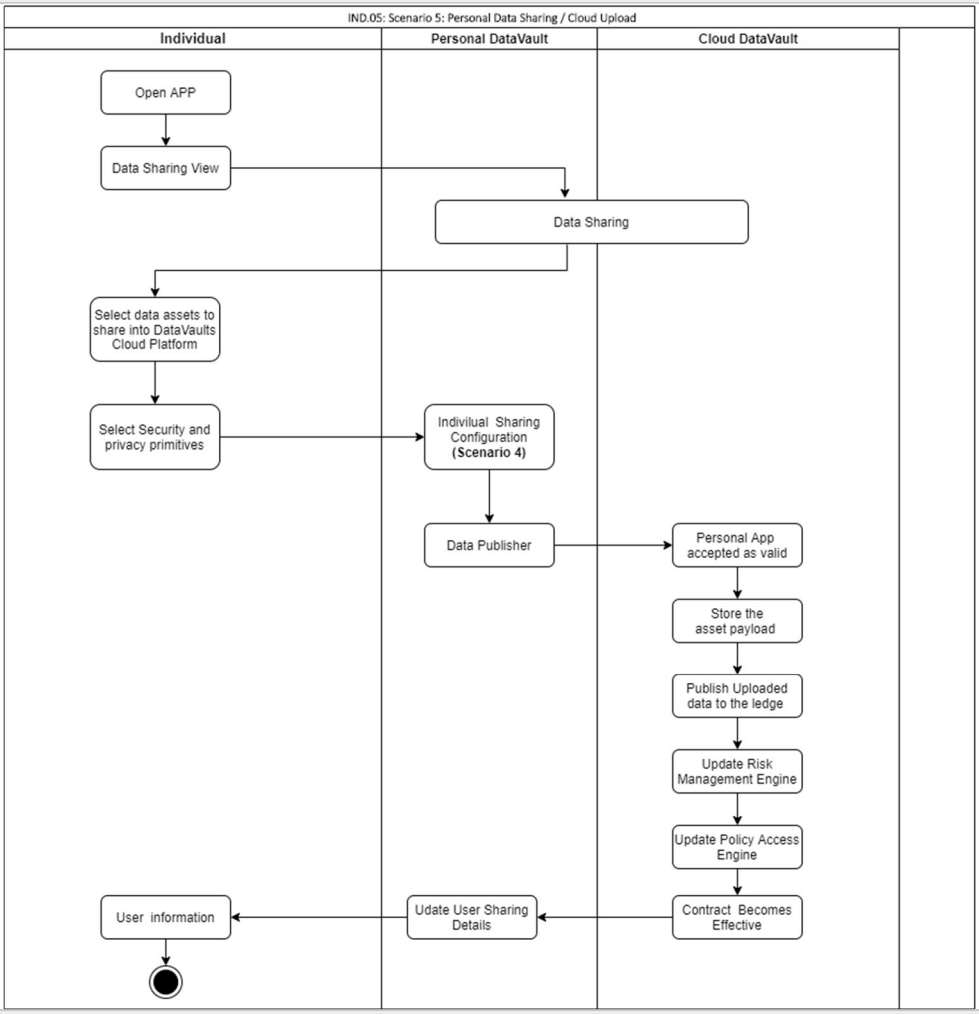
Triggers	● An Individual is willing to share their personal data with DataVaults and wants to configure parameters about data sharing using DataVaults Personal App. ● A predefined request for sharing data arrives and the sharing configurations is preloaded and presented to the Individual.
Workflow	
Scenario Sequence	1. The Individual opens/logs into DataVaults Personal App and visits/navigates to the area dedicated to data sharing. 2. The Individual may choose an already stored data sharing configuration template or initiate a new configuration. In case of a template, the

	operations 3-10 below are pre-loaded and s/he needs to go through them and approve them. 3. The Individual selects one or more categories of assets (e.g. health related data) and/or one or more assets (e.g. heart rate as recorded by his/her smart watch) s/he wants to share with Data Seekers through the DataVaults Cloud Platform. 4. The Individual selects which part of the data will be shared with DataVaults Cloud Platform (i.e. all social media activity minus geo-location information). 5. The Individual selects if the data will be shared in full or/and anonymized, as there is the option to turn off privacy features for data sharing, to share eponymous data. 6. In case of anonymized sharing, the Individual selects, prior to sharing and uploading data to DataVaults Cloud Platform, if they want their data to be shared as personal anonymized data (i.e. Digital Twin) or as grouped anonymized data (i.e. to become part of a Persona Group). (Anonymiser) and (Identities Wallet). 7. The Individual indicates different access policies on the assets to dictate how these (or their derivatives) can be searched, requested (directly provided by the platform or after user consent), used, for how long, by which types of Data Seekers (i.e. only public organisations, but not private companies, for one year from now) (Access Policy Editor). 8. The Individual sets a price tag on the assets to be shared, with each price tag linked to one of the policies chosen, allowing him/her to share assets with different pricings schemes per policy. 9. The Individual selects if he would like to share the data on regular intervals, and for how long, or if this a one-time only sharing action. For specific data types that have an expiration/obsolete date, the Individual might choose to auto-share the renewed data as soon as it is fetched, removing the old one (link to Scenario 6). 10. Based on the Individual's selections, the DataVaults Personal App informs the Individual about the projected privacy risk exposure (Privacy Metrics Dashboard). 11. The Individual makes their final decision and provides a description for the data assets to share with DataVaults Cloud Platform (i.e. "my social media activity for public"), providing also a set of metadata. 12. The Individual can save the sharing configuration to re-use it with the same or any other dataset in the future.
User Benefits	The user is able to configure a multitude of parameters (incl. projected earnings) related to sharing his/her own personal data with Data Seekers through the DataVaults Cloud Platform.
Challenges	● The data selection process should be made very user friendly to be operated over portable devices. ● On-the-fly accurate calculation of risk exposure is a difficult task. ● Setting the appropriate value tags is essential for data to attract interest.
GDPR Issues	● Transparency needs to be ensured (see comment to Scenario 1). It is key that all necessary information is provided before the choice of the data sharing configuration template (SS 3-10), with special attention in case of one or more categories of assets (e.g. health related data). Information

	should also include in advance the privacy risk exposure (Privacy Metrics Dashboard) (see Scenario 3) and later on, in case of need, a future warning of the increased level of risk. Clear explanation of the differences between Digital Twin and Persona Group options should be explained (SS 6), focusing on main features rather than on technical details. ● The templates for the configuration of parameters about data sharing should rely on the data minimisation principle, so that, by default, only personal data which are necessary for each specific purpose of the processing are processed.
Ethical Issues	● Any form of discrimination and stigmatization needs to be avoided, for instance averting to differentiate the price/reward according to the level of anonymisation chosen or otherwise pushing towards the turning off of privacy features for data sharing (SS 5). ● Non-monetary compensation forms need also to be considered in SS 8.
Privacy Issues	● User and Data Privacy: Privacy features, selected by the users, for sharing their data should be in line with the privacy preferences that were selected by the users when granting authorization to the DataVaults Personal App for data collection (Scenario 2). This entails both the provision (SS 6) of only those appropriate anonymization data mechanisms (e.g., Digital Twins, User Personas, Direct Anonymous Attestation), based on the defined user privacy features of interest, and the verification that the implementation of these mechanisms can achieve the required privacy preservation (as acceptable threshold calculated by the Datavaults Risk Assessment component). ● Fairness and User Data trading Privacy: The allocation of a price tag (SS 8), by the users, to the data assets to be shared should not allow: (i) any user privacy breach by enabling any malicious user or Observer to link trading transaction values to specific user identities (no price tags that can limit the user anonymity set), and (ii) misbehaving Data Seekers to exploit the trading mechanisms for increasing their utility without making the requested contributions/payments. This will be achieved through the integration of appropriate homomorphic encryption mechanisms and individual Blockchain User Wallets for the security of data trading transactions.
Security Issues	● Authorization and Attributed-based Access Control: The representation and deployment of the (defined) user data sharing policies will be achieved through the use of smart contracts and Attribute-based Access Control (ABAC) mechanisms integrated to the DataVaults framework. Thus, users, when selecting/defining the data sharing configuration policy that will underpin their future data trading transactions, should be presented with the correct interpretation of such policies to the attributes, that Data Seekers need to exhibit for being able to procure data of interest, as evidence of the correct access control policies definition. This will be achieved through the DataVaults ABE Trusted Component that will enable the secure smart contract computation and verification functionalities by the users. ● Non-repudiation of User's Data Sharing Configuration: Users should not be able to dispute (during a later data trading transaction) that a data sharing configuration policy (accompanying a specific data set) was selected/defined by them. This will be achieved through the use of appropriate signature mechanisms.

	• Operational Correctness of the execution of the selected anonymization techniques so as to provide verifiable evidence that these mechanisms were employed correctly and achieve the desired user privacy protection. This will be achieved through the DataVaults Data Anonymizer component and the trusted execution of user-controlled anonymous attestation functionalities.

3.2.5 Scenario 5: Personal Data Sharing / Cloud Upload

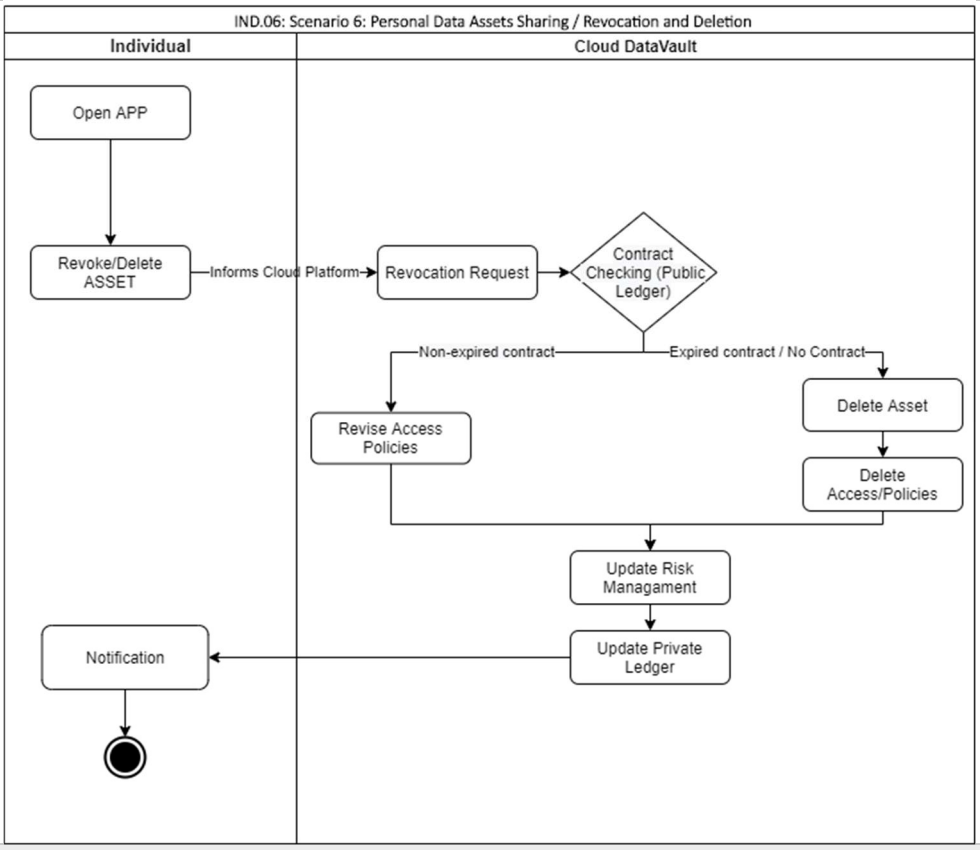
Scenario ID	IND.05
Scenario Title	Personal Data Sharing / Cloud Upload
Scenario Actors	Individual, DataVaults Personal App, DataVaults Cloud Platform
Overview	This scenario describes all actions that take place at the moment when an Individual chooses to upload the data assets to the DataVaults Cloud Platform.
Triggers	An Individual is willing to share their personal data with DataVaults, has already set sharing configuration parameters (through Scenario 4: Personal Data Assets Sharing Configuration) and is ready to upload his/her data to the DataVaults Cloud Platform.
Workflow	
Scenario Sequence	1. The Individual selects to upload his/her data assets to DataVaults Cloud Platform. 2. The DataVaults Personal App selects the appropriate security and privacy primitives to be attested by the DataVaults Cloud Platform.

	3. The DataVaults Personal App sends data to DataVaults Cloud Platform (Data Publisher) while preserving sharing configuration set by the Individual (anonymisation preference, access policy rules, asset value pricing, asset description, etc.). 4. The DataVaults Personal App is attested by the DataVaults Cloud Platform and the payload is accepted as valid. 5. The asset payload is stored securely in the DataVaults Cloud Platform. 6. A record of the data upload is pushed to the ledger. 7. The Risk Management Engine is updated using the information of the transaction to update the risk exposure metrics of the Individual. 8. The Policy Access Engine is updated with the configuration relevant to this transaction. 9. The Individual is notified about the success of the operation and the transaction is visible in his/her DataVaults Personal App, also being provided with updated metrics regarding his/her risk exposure.
User Benefits	The Individual has successfully uploaded his/her personal data to DataVaults Cloud Platform.
Challenges	● Attestation between the Personal App and the Cloud Platform needs to be robust and fast. ● Attestation between the Personal App and the Cloud Platform might limit the available hardware that can be used. ● Ledger operations needs to be fast.
GDPR Issues	● The data owners should be informed on the specific Data Seekers and the purpose of their further processing (in case of not anonymized data), in order to be able to provide consent.
Ethical Issues	● It is key to properly face the management of privacy / utility trade-offs, also as regards Data Seekers and Secondary Use, preserving the utility for data analysis, as well as the control of the Individual upon his/her personal data. ● A layered approach to consent management could be explored, providing all the necessary information step by step, where it is not essential that the first layer of information contains all the details of the processing or further processing. It should be also investigated which information needs to be given to the individual in which layer and if, in some advanced layer, the implicit consent is adequate (though not in case of special categories of personal data listed in Art. 9 GDPR).
Privacy Issues	• User Anonymity and Unlinkability: User actions regarding the uploading of data assets (SS 3), to the DataVaults Cloud Platform, should not breach the anonymity level that was defined in the user privacy policy. This entails that no specific data upload action should be linked back to the original initiator without breaching the non-repudiation requirement of the transaction. Linkability should only be checked and achieved by the host trusted component (as part of the individual's Blockchain Wallet) for the successful compensation of the users when part of their data is procured by Data Seekers (thus, achieving privacy-preserving data trading and user compensation). • Data Privacy: The data asset payload that is sent to the DataVaults Cloud Platform, should be stored securely (SS 5) so that no adversary can breach its confidentiality that can then also lead to the exposure of the user's id that is the owner of this data bundle. This entails that no malicious user or component can learn the secret information (attribute-based key) used

	for the attributed-based encryption of this data asset. This will be achieved through the integration of the DataVaults ABE Trusted Component that will be responsible for the secure (attested) management of all secret information required for any data upload/sharing action. • Data Integrity: The shared data asset payloads must be protected with appropriate controls to ensure their integrity when uploaded by the DataVaults Personal App to the cloud platform.
Security Issues	• Confidentiality of the data uploaded to the DataVaults Cloud Platform (SS 3) by providing a secure communication channel between the DataVaults Personal App and the Data Publisher. This will be achieved through the use of strong crypto primitives (i.e., AES) and attestation mechanisms. Key management of the respective secrets will be achieved through the host trusted component. • Non-repudiation of User's Data Upload: Users should not be able to dispute the secure uploading of any data bundle to the DataVaults Cloud Platform. Thus, verifiable evidence should be kept for the auditability of all such transactions in the distributed ledger. • Smart Contract Verification: The Policy Access Engine should provide verifiable evidence that the correct smart contract, corresponding to the data sharing configuration and access control policy linked to the user's data asset payloads, is correctly calculated and pushed to the ledger. Smart contracts, in the private ledger, will be used to process information about the access policies (per data asset), thus, they should adhere to the data sharing configurations defined by the users (Smart Contract Trusted Control Services). • Ledger Security: As in Scenario 3, all data upload transactions that are registered in the ledger, should not be tampered with. This will be achieved through the DataVaults Blockchain verification and validation mechanisms of all recorded transaction blocks.

3.2.6 Scenario 6: Personal Data Assets Sharing Revocation and Deletion

Scenario ID	IND.06
Scenario Title	Personal Data Assets Sharing Revocation and Deletion
Scenario Actors	Individual, DataVaults Personal App, DataVaults Cloud Platform
Overview	This scenario describes all the actions that an Individual performs to revoke the sharing of his/her data assets with Data Seekers through the DataVaults Cloud Platform and deletion
Triggers	An Individual wishes to revoke the personal data assets previously shared with Data Seekers through the DataVaults Cloud Platform.

Workflow	
Scenario Sequence	1. The Individual (or the DataVaults Personal App in case a data asset expires), selects a data asset already shared over the DataVaults Cloud Platform (e.g. heart rate as recorded by his/her smart watch, a certificate, etc.) and chooses to revoke sharing and delete the data asset completely from the Cloud Based Platform. 2. The DataVaults Personal App informs the DataVaults Cloud Platform about the revocation request of the shared data asset. 3. The DataVaults Cloud Platform checks if the data asset under revocation is acquired by one or more Data Seekers and the terms accompanied in the contracts of these transactions: a. in case the data asset is part of a non-expired contract, revocation is not possible, but the asset will become inaccessible from any other Data Seekers. b. In case the data asset is not part of a data contract, then the asset is deleted from the DataVaults Cloud Platform. 4. The DataVaults Cloud Platform deletes access policies related to the data asset under revocation. 5. The Risk Management Engine is updated using the information of the transaction to update the risk exposure metrics of the Individual. 6. The Private Ledger is updated with a record of the asset revocation from the user. 7. The Individual is notified about the success of the operation, also being provided with updated metrics regarding his/her risk exposure (Privacy Metrics Dashboard).
User Benefits	● The Individual is able to revoke access to his/her shared data assets. ● Data Seekers who have a valid contract allowing the indefinite use of the asset, can keep on using it.

Challenges	• Revoking assets used already in Personas might be difficult • Revoking assets that are part of active contracts needs a mechanism to check contract expiry dates and resolve other clauses included in the contact
GDPR Issues	• Further investigation needs to be conducted regarding the impossibility to revoke consent in case the data asset is part of a non-expired contract (as well as the position of Data Seekers having a valid contract in place allowing the indefinite use of the asset) (SS 3 a). Likewise, this additional research needs to be done in case the revoked data assets were used for building Personas. • The withdrawal is exercised for the future without retroactive effect: this implies that 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 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), the other legitimate grounds for personal data processing and the limits to their applicability and the switch from one legal basis to another, as well as the interest of the Data Seekers. • The Privacy Protection Goals method, merged with the Privacy-by-Design and-by-Default approach, which is at the core of the DataVaults Ethical Policy (as described in DataVaults D2.1 and D9.2) will drive such future work. The legitimacy and fairness of technologies will 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.
Ethical Issues	• The same as GDPR issues.
Privacy Issues	• Forward User and Data Privacy: The revocation of data sharing credentials should not affect the unlinkability of any other uploaded data by the same user. More specifically, during the revocation of the sharing credentials of the data assets to be deleted (access control policy and unique encryption key based on the employed ABE mechanism) (SS 4), it is imperative that no link to the user ID (that owns this data) or the other data assets, belonging to the same user, can be inferred. User and (remaining) data privacy should be kept as well as the secrecy of all other attribute-based keys that have been generated even if the same set of attributes need to be exhibited by Data Seekers for getting access to the data resources. This will be achieved by the appropriate key generation process to be executed by the DataVaults ABE Trusted Component. • Unlinkability of created User Personas: The same, as above, should also be the case for the deletion of the selected data assets from any created user personas. When specific data assets are deleted, from created user personas, unlinkability to the user ids from whom (obfuscated) data are also included in these personas, should be preserved. Especially considering that the creation of such personas is based on the obfuscation and merging of data originating from multiple users with similar characteristics, it is of paramount importance to preserve their privacy.

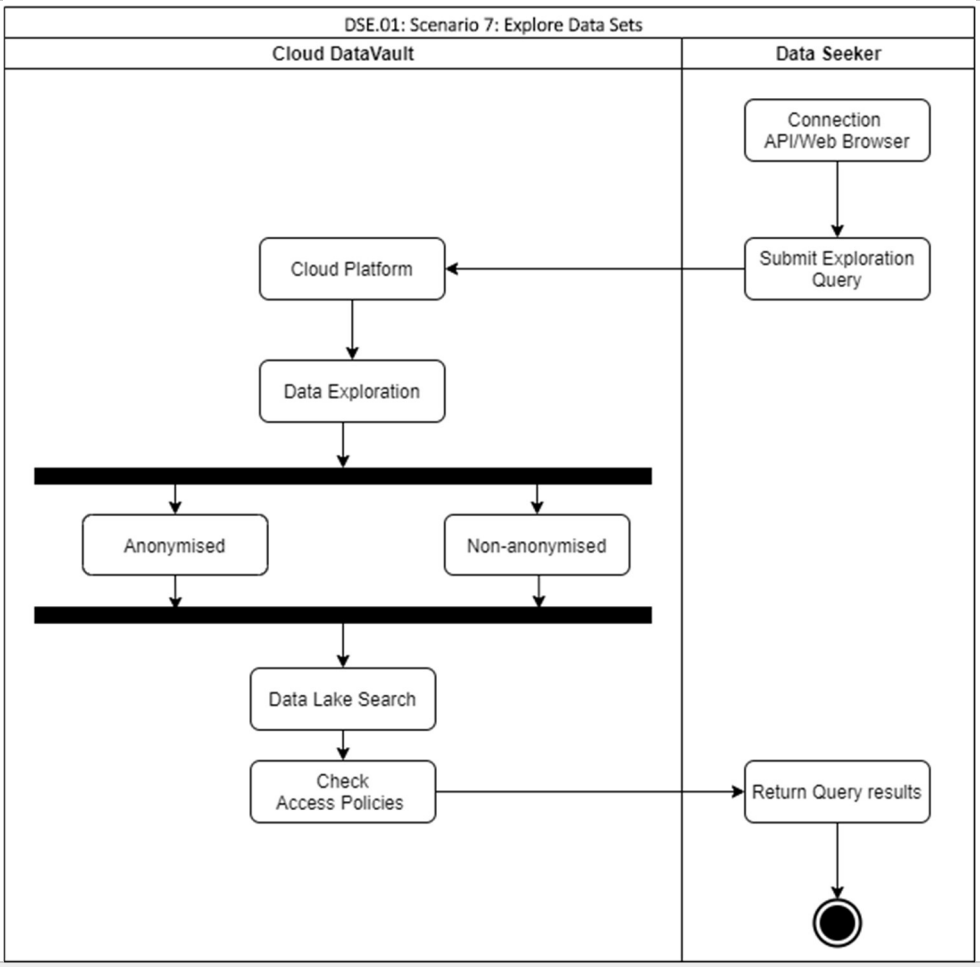
Security Issues	• Secure Access Control Policy Deletion: The removal of the access control policy, associated with the data asset selected for deletion, should be authorized only by an authenticated user (data owner) account. Furthermore, the smart contract that has been compiled for enforcing the data sharing policy must be securely updated to correctly reflect that the target data asset is no longer accessible from any other Data Seekers; either by deploying a new contract or by dynamically updating the already existing contract (see D2.2 for more information). • Secure Deletion of Data Pointer on the Ledger: All pointers that have been created to accompany the data (securely) stored on the DataVaults Cloud Platform – that are used for the efficient identification of data assets in the Encrypted Searchable Data Lake – should also be safely and securely disposed. If such pointers are pushed in the ledger then verifiable evidence should be provided that the respective DataVaults Cloud Platform Asset Indexing (Operation V.3) is no longer active. • Non-repudiation of User's Data Deletion: Users should not be able to dispute the deletion request of any data bundle to the DataVaults Cloud Platform. Thus, verifiable evidence should be kept for the auditability of all such transactions in the distributed ledger.

3.3 SCENARIOS DRIVEN BY DATA SEEKERS

In this Section, four scenarios that describe the interactions of the Data Seekers with DataVaults are presented. These scenarios guide the reader through the steps and other details of various operations driven by a Data Seeker, such as the exploration of personal data available through the DataVaults Cloud, the acquisition of data and compensation of the Individuals, as well as the analysis and visualisation of data in the Data Seeker's Vault.

3.3.1 Scenario 7: Explore Data Assets

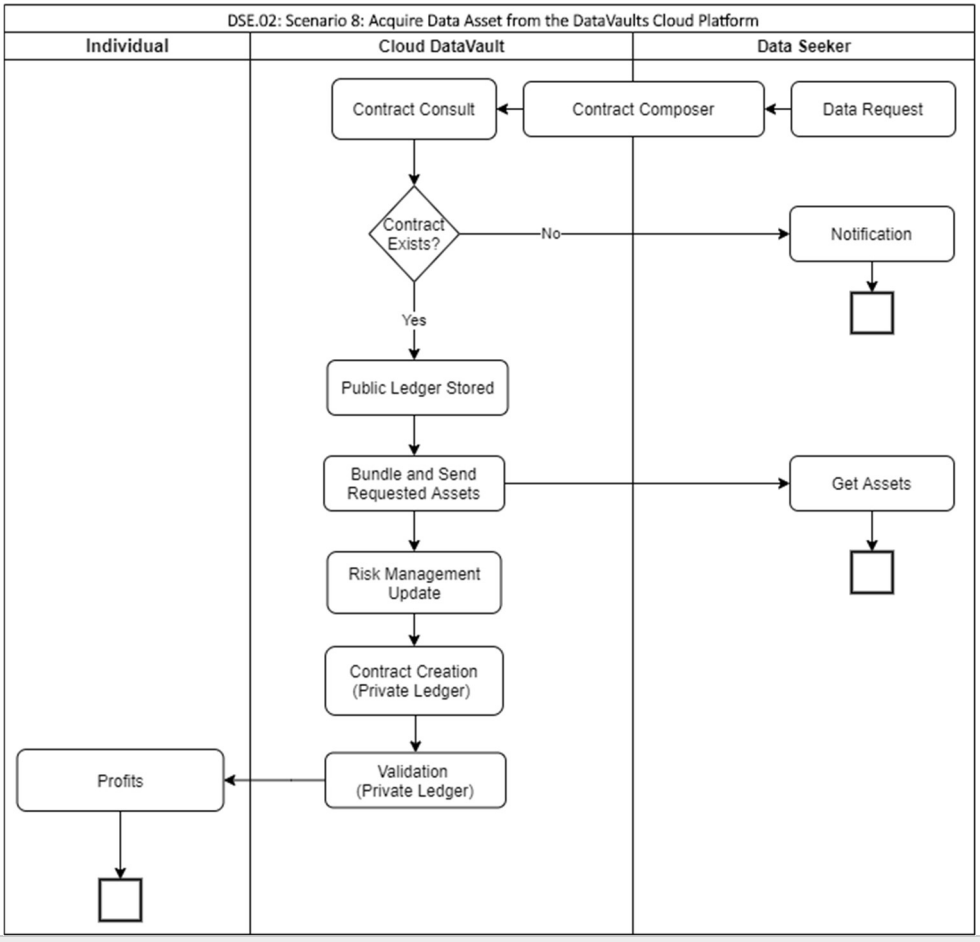
Scenario ID	DSE.01
Scenario Title	Explore Data Assets
Scenario Actors	Data Seeker, DataVaults Cloud Platform
Overview	A Data Seeker wants to explore data assets that are made available over the DataVaults Cloud Platform. Specifically, the Data Seeker connects to DataVaults Cloud Platform through different interfaces and explores data using search mechanisms such as the DataVaults Data Lake, in order to locate data which s/he is interested in.
Triggers	A Data Seeker connects to DataVaults Cloud Platform with the intention to explore anonymous data.

Workflow	
Scenario Sequence	1. The Data Seeker connects to DataVaults Cloud Platform: a. Connection via an API through a third-party application. b. Connection directly through web/browser-based access. 2. The Data Seeker explores, by authoring a search query: a. anonymous data. In this case the Data Seeker can choose if s/he wishes to search for collective/aggregated anonymised data of a set of Individuals (grouped under a "Persona") or individual anonymised data (Digital Twin). b. unmasked (eponymous) data from Individuals. In this case one or more Individuals have already shared their personal non-anonymised data with the DataVaults Cloud Platform. 3. The DataVaults Cloud Platform searches over the Encrypted Data Lake, as well as over unencrypted data stored in the Cloud Platform 4. The DataVaults Cloud Platform checks the Access Policies which are in place related to the requested data and offers, accordingly, a snippet of the results and a pointer of where the data are, alongside with a set of metadata that describe them.
User Benefits	The Data Seeker is offered: ● Fundamental search functionality, allowing the Data Seeker to look for personal data which interest him/her before s/he proceeds to acquire them. ● Opportunity to search through unmasked (eponymous) personal data, anonymous personal data, encrypted or non-encrypted.

Challenges	● Searchable Encryption concerns technologies that are not highly mature to display results at the level a seeker wants to.
GDPR Issues	● In case of personal data (SS 2 b), the same considerations as in the Scenarios driven by Individuals apply.
Ethical Issues	● In case of personal data (SS 2 b), the same considerations as in the Scenarios driven by Individuals apply.
Privacy Issues	● User Anonymity and Unlinkability: Data Seeker actions regarding the exploration of data assets (SS 2) should not breach the anonymity level that was defined in the privacy policy configured by the data owner. This entails that no specific data exploration action should be linked back to the original data owner. ● User Privacy: The data asset payload that is sent to the Data Seeker, should not reveal any identifying information about the owner(s) of the data that can lead to the exposure of the user's id. This also entails that no link should be feasible to other data bundles that have been uploaded, to the DataVaults Cloud framework, by the same user(s). ● Data Integrity: The explored data asset payloads must be protected with appropriate controls to ensure their integrity when shared by the DataVaults Cloud Platform to the requested Data Seekers.
Security Issues	● Data Seeker Attributes Correctness: When the DataVaults Cloud Platform checks the access policies that accompany the requested data (SS 4), it has to verify that the Data Seeker exhibits the correct list of attributes required for having access to these data resources. This will be achieved through the integration of appropriate Attributed-based Encryption mechanisms and the DataVaults ABE Trusted Component. ● Smart Contract Verification: The Policy Access Engine should provide verifiable evidence that the correct smart contract, corresponding to the data sharing configuration and access control policy linked to the requested data asset payloads, is correctly executed by the Data Seekers. Smart contracts will be used to process information about the access policies (per data asset), thus, they should be correctly enforced prior to sharing information about collected data (Smart Contract Trusted Control Services). ● Non-repudiation of Data Seeker's Data Exploration: Data Seekers should not be able to dispute the secure exploration of any data bundle to the DataVaults Cloud Platform. Thus, verifiable evidence should be kept for the auditability of all such transactions in the distributed ledger. ● Ledger Security: As in Scenarios 3 and 5, all data sharing transactions that are registered in the ledger, should not be tampered with. This will be achieved through the DataVaults Blockchain verification and validation mechanisms of all recorded transaction blocks.

3.3.2 Scenario 8: Acquire Data Assets from the DataVaults Cloud Platform

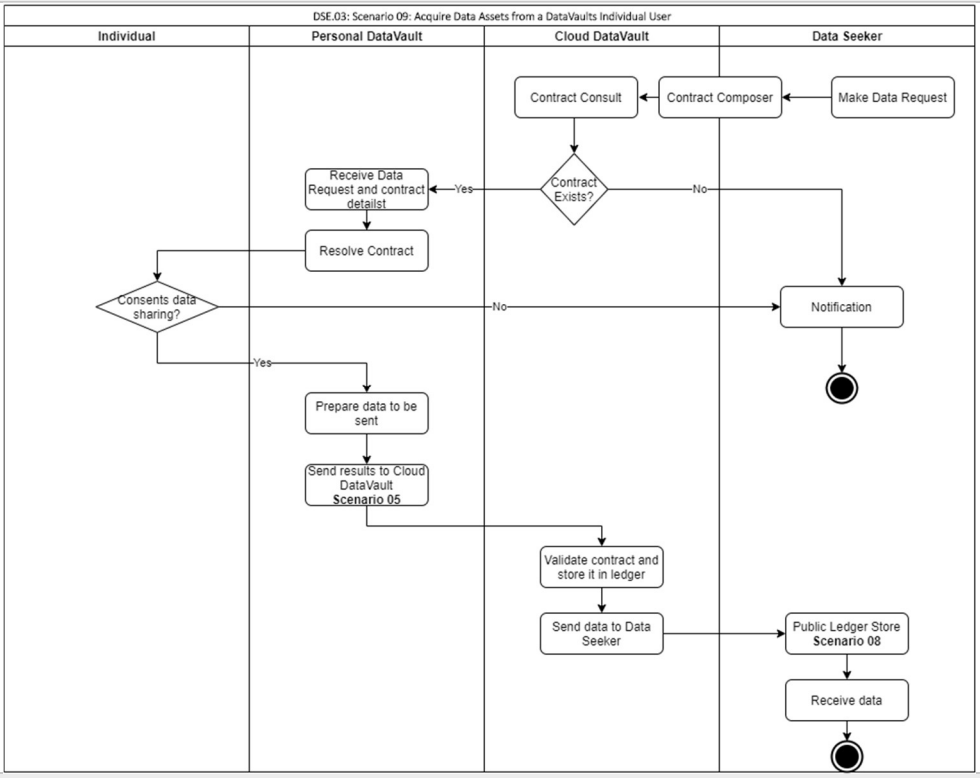
Scenario ID	DSE.02
Scenario Title	Acquire Data Assets from the DataVaults Cloud Platform
Scenario Actors	Data Seeker, DataVaults Cloud Platform
Overview	A Data Seeker wants to acquire data which have been made available from Individuals through the DataVaults platform. Specifically, the Data Seeker requests data assets s/he has already located (e.g. through exploration) and offers the listed

	compensation for those by issuing a smart contract. Following this, the DataVaults Cloud Platform resolves the request, validates the contract and provides the asset(s) to the Data Seeker, and transfers the relevant compensation to the Individuals.
Triggers	A Data Seeker is already connected and wants to acquire data from Individuals through the DataVaults platform. The data is assumed to be already shared with the DataVaults platform by Individuals under a certain configuration (data asset type selection, anonymisation mechanisms used, access policies design, asset value description, etc.), for more details see Scenario 3: Personal Data Assets Sharing Configuration.
Workflow	
Scenario Sequence	1. The Data Seeker makes a Data Request via the DataVaults Platform (Data Request Service) to acquire some assets he has spotted 2. The DataVaults Cloud Platform (Contract Composer) issues a smart contract between the Data Seeker and the DataVaults Cloud Platform for the selected assets. 3. The DataVaults Cloud Platform consults whether access can be provided to these assets for the specific profile of the Data Seeker a. [*alt-flow 1*] If access is not provided, then the smart contract above is not validated and the Data Seeker is notified about the failure of the operation. **The Usage Scenario ends here.** 4. In case the assets are available, the value is transferred from the Data Seekers wallet to the Cloud Platform and the smart contract is validated and stored in the Public ledger 5. The assets are provided to the Data Seeker unencrypted

	6. The Risk Management Engine is updated using the information of the transaction to update the risk exposure metrics of the Individual. 7. A smart contract in the Private Ledger is created and validated between the DataVaults Cloud Platform and the Individual whose assets have been sold, and the appropriate earnings are transferred to the Individual's wallet
User Benefits	● The Data Seeker acquires personal data from one or more Individuals. ● The Individual is able to sell assets using predefined preferences, without being asked all the time
Challenges	● Validating the contract on the 2 ledgers might prove resource heavy ● Decryption of the assets might require a complex sequence of key exchanges
GDPR Issues	● The same considerations described in the Scenarios driven by Individuals apply, for instance as regards transparency, information on risk exposure and on Data Seekers' purpose of the processing, informed consent and compensation mechanisms. All the data subjects' rights laid down in GDPR must be respected in the whole flow. ● Though an Individual can benefit from predefined selling preferences, nevertheless it is necessary to have his/her specific consent to any data sharing, seeking to identify and implement modalities to avoid the "consent fatigue".
Ethical Issues	● The same considerations described in the Scenarios driven by Individuals apply.
Privacy Issues	● The same considerations described in Scenario 7 apply driven by Data Seekers when they want to explore specific data asset payloads. ● Fairness and User Data trading Privacy: Data seekers should not be able to exploit the trading mechanisms for neither breaching user privacy by enabling them to link trading transaction values to specific user identities, nor increasing their utility without making the requested contributions/payments. This will be achieved through the integration of appropriate homomorphic encryption mechanisms and individual Blockchain User Wallets for the security of data trading transactions.
Security Issues	● The same considerations described in Scenario 7 apply driven by Data Seekers when they want to explore specific data asset payloads.

3.3.3 Scenario 9: Acquire Data Assets from a DataVaults Individual User

Scenario ID	DSE.03
Scenario Title	Acquire Data Assets from a DataVaults Individual User
Scenario Actors	Data Seeker, DataVaults Cloud Platform, DataVaults Personal App, Individual
Overview	A Data Seeker is already connected and wants to acquire data directly from Individuals through the DataVaults platform. Specifically, the Data Seeker requests the data of one or more Individuals through DataVaults Cloud Platform and proposes a fair compensation for those by issuing a smart contract. Following this, the DataVaults Cloud Platform forwards the request to DataVaults Personal App. The Individual is informed and chooses to accept the compensation, or not. Following this, the DataVaults Personal App resolves the request, either by validating the contract, picking the data and providing them back to the Data Seeker and instructing the DataVaults Cloud Platform to perform the necessary compensation transfer actions, or by rejecting the contract.

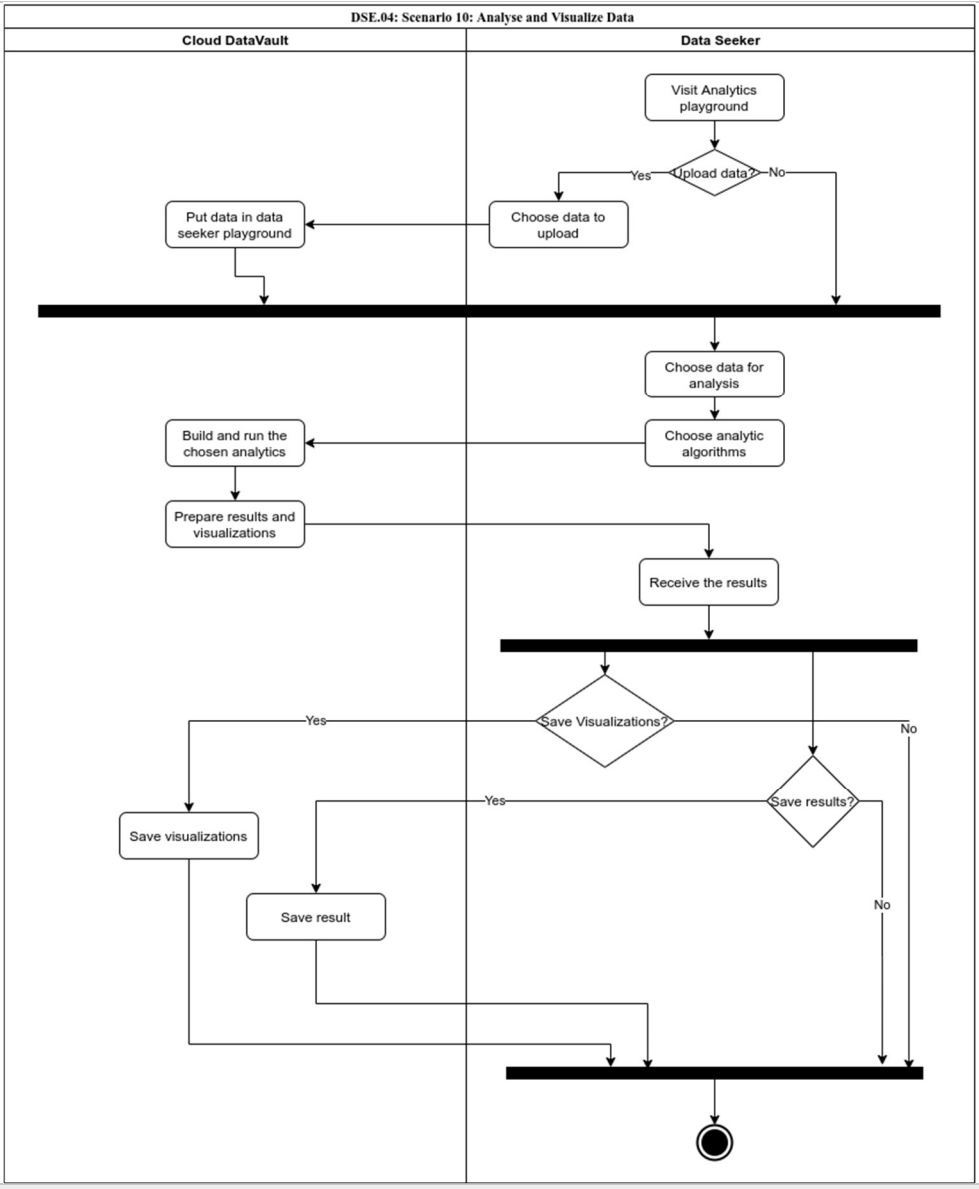
Triggers	A Data Seeker wants to acquire data from Individuals through the DataVaults platform. The data is assumed to be at the side of the Individual who has not yet shared them with the DataVaults platform.
Workflow	
Scenario Sequence	1. The Data Seeker makes a Data Request via the DataVaults Platform (Data Request Service) to acquire some assets from an Individual he has spotted which are marked as not shared over the platform, or some other data assets that he believes the Individual can provide but are not listed in the platform (like questionnaire replies). 2. The DataVaults Cloud Platform (Contract Composer) issues a smart contract for the selected data to be stored on the Public Ledger 3. The Data Request and smart contract details are forwarded to the Individual's DataVaults Personal App. 4. The Individual's DataVaults Personal App resolves the Data Request (Data Request Service Resolver) and gets the defined configuration for sharing these data from smart contract. 5. The Individual accepts the offered compensation and verifies it through the DataVaults Personal App. a. [*alt-flow 1*] The Individual does not accept the offered compensation; the Data Seeker is informed about the rejection. **The Usage Scenario ends here.** 6. The data are picked from Individual's side (Data Picker) or are directly uploaded/edited by the Individual to the Data Picker (in case of data not already on the App such as questionnaire data) 7. A private smart contract is created between the Individual and the DataVaults Cloud Platform for the data transfer 8. The assets are transferred to the DataVaults Cloud Platform in a secure manner (see scenario IND.05)

	9. The smart contract is validated by the DataVaults Cloud Platform and is stored on the Ledger, and the value is transferred from the Data Seeker to the DataVaults Cloud platform. The private smart contract (between the Individual and the Cloud Platform) is validated and stored in the private ledger, transferring the earnings from the Cloud platform to the Individuals' wallet 10. The data are provided to the Data Seeker through the DataVaults Cloud Platform
User Benefits	● The Data Seeker acquires personal data from one or more Individuals. ● In this usage scenario the Individual is able to negotiate the compensation for his/her personal data.
Challenges	● Selection of Individuals for data acquiring in case these are grouped under a Persona, does not provide Guarantees about the quality of data requested ● Back-and forth negotiation regarding the compensation means that a dedicated area in the UI should be created for this reason. ● Passing the data through the DataVaults Cloud Platform would need to establish a secure communication channel to fetch and store these data at the Data Seekers side
GDPR Issues	● Request for selecting such data should be made without revealing the real identity of Individuals. ● The Individual must be safeguard against possible negative effects of his/her refusal (such as being discarded in future queries of that or other Data Seekers). ● The same issues described in Scenarios driven by Individuals apply.
Ethical Issues	● The same issues described in Scenarios driven by Individuals apply.
Privacy Issues	● User Forward Privacy and Unlinkability: The data asset sent to the Data Seeker should not reveal any identifying information about the owner of the data that can lead to the exposure of the user's id. This also entails that no link should be feasible to other data bundles of the same user. Data assets can be seen to belong to pseudonyms (i.e., user pseudo-identities) so that appropriate compensation can be transferred to the users (linked to those pseudonyms) but no other identification should be feasible. In this context, the DataVaults Cloud platform should act as the intermediate. ● Data Integrity: The shared data asset payloads must be protected with appropriate controls to ensure their integrity when uploaded by the DataVaults Personal App to the cloud platform.
Security Issues	● Confidentiality of the data uploaded to the DataVaults Cloud Platform (by the DataVaults Personal App) and then forwarded to the requested Data Seeker by providing secure communication channels between the communicating parties. This will be achieved through the use of strong crypto primitives (i.e., AES) and attestation mechanisms. It is imperative that secrets\keys, created for each communication channel, are unique so that there is no need to share the user's individual key (used for the secure transfer of data to the DataVaults Cloud platform – SS 6) with the Data Seeker that will then require a possible update of all other user keys. ● Non-repudiation of Data Seeker's & User's Actions: Both entities should not be able to dispute the actions related to a data trading transaction: Data Seekers should not be able to dispute the secure acquisition of any data bundle that was shared by specific users. Thus, verifiable evidence

	should be kept for the auditability of all such transactions in the distributed ledger. ● Transaction Verification: Both Users and Data Seekers should be able to verify the correctness of all (related) trading transactions, presented by the DataVaults framework, so as to have strong guarantees that these were executed correctly based on the agreed policies enforced through the smart contracts. This pertains to both the level of data sharing, included in the transaction, as well as the correct user compensation. It will be achieved through the DataVaults distributed ledgers and the provided management and access interfaces; i.e., the ID (or link) of the transaction should be also presented to the user so as s/he can check and verify the respective transaction block stored in the DataVaults private ledger. ● Ledger Security: As in the previous scenarios, all data sharing transactions that are registered in the ledger, should not be tampered with. This will be achieved through the DataVaults Blockchain verification and validation mechanisms of all recorded transaction blocks

3.3.4 Scenario 10: Analyse and Visualise Data

Scenario ID	DSE.04
Scenario Title	Analyse and Visualise Data
Scenario Actors	Data Seeker, DataVaults Cloud Platform
Overview	A Data Seeker uses DataVaults Cloud Platform to analyse personal data s/he has already acquired through DataVaults Cloud Platform, possibly combine them with personal data s/he has already in his/her possession and visualise the results in order to extract insights.
Triggers	A Data Seeker wants to analyse and visualise data using DataVaults Cloud Platform.

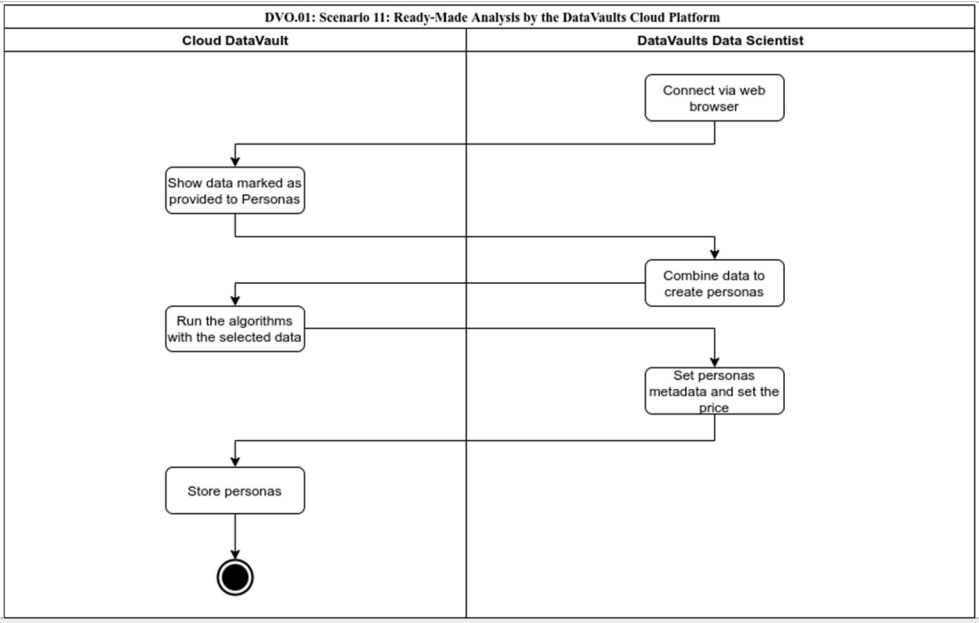
Workflow	
Scenario Sequence	1. The Data Seeker visits the "Analytics Playground" area of the DataVaults Cloud Platform. 2. The Data Seeker can upload other data s/he already has in his/her possession at his/her side. 3. The Data Seeker can choose which data of those acquired through DataVaults and the ones he uploaded he would like to consider for the analysis 4. The Data Seeker is able to construct analyses and apply different algorithms and statistical methods to the assets selected (Secure Analytics Playground). 5. The Data Seeker visualises the results of the analysis (Visualization Dashboard). 6. The Data Seeker can save or export his/her queries for future re-use. 7. The Data Seeker can save export his/her visualisation for future re-use.
User Benefits	● The Data Seeker is able to use a multitude of algorithms and statistical methods for personal data analysis. ● The Data Seeker can combine personal data s/he already possesses with personal data of interest acquired through the DataVaults Cloud Platform.

	● The Data Seeker can save his/her queries and visualisations for future re-use (e.g. after the acquisition of more personal data of interest).
Challenges	● Data Analytics need to be performed with an Intuitive UIs (such as Notebooks), covering the requirements of Data Analysts ● Data Analytics might require large infrastructure resources for selected algorithms
GDPR Issues	● In case the Data Seeker combines personal data s/he already possesses with personal data of interest acquired through the DataVaults Cloud Platform, s/he must have a legitimate ground also for the data processing of such dataset already possessed. Likewise, GDPR requirements apply also in relation to it.
Ethical Issues	N/A
Privacy Issues	● User Anonymity: Data Seeker actions regarding any requested analysis (on their acquired data) should not breach the anonymity/privacy level of the initial user data owners. This entails that no personal information should be outputted by the analysis performed by DataVaults that can be used by an Observer to infer private user information or any link to their identities. This will be achieved through the employment of appropriate differential privacy algorithms on the data yielded from the analysis tasks.
Security Issues	● Confidentiality and Authentication: Only the correct Data Seeker, as the data procurer, should be able to have access to this DataVaults area and perform data analysis tasks. This will be achieved through the appropriate Data Seeker authentication and access control mechanisms, employed by the DataVaults Cloud platform, when logging in to the framework. Certificates will be leveraged through the host trusted component. ● Non repudiation: Data Seekers should not be able to dispute a data analysis action that was initiated by them that requires some form of data processing. This will be achieved through the use of appropriate signature mechanisms.

3.4 SCENARIOS DRIVEN BY THE DATAVAULTS DATA SCIENTIST

3.4.1 Scenario 11: Ready-Made Analysis by the DataVaults Cloud Platform

Scenario ID	DVO.01
Scenario Title	Create Ready-Made Analysis of Assets available in the DataVaults Cloud Platform.
Scenario Actors	DataVaults Data Scientist, DataVaults Cloud Platform
Overview	A DataVaults Data Scientist connects to DataVaults Cloud Platform in order to analyse data and create ready-made analyses for others (e.g. Data Seekers) to find available upon connection.
Triggers	A DataVaults Data Scientist connects to DataVaults Cloud Platform with the intention to perform one of his/her tasks.

Workflow	
Scenario Sequence	1. The DataVaults Data Scientist connects to DataVaults Cloud directly through web/browser-based access. 2. The DataVaults Data Scientist explores anonymous data shared by Individuals with the DataVaults Cloud Platform, which are marked by individuals as assets they wish to provide to Personas. 3. The DataVaults Data Scientist combines personal data from his/her executed queries in order to group them and create Personas. 4. The DataVaults Data Scientist saves his/her queries in the DataVaults Cloud Platform for future re-use by other users, providing some metadata on the created Persona and setting the price for it, guided by the price that the included users requested.
User Benefits	● The Data Seekers are able to get meaningful Personas created by a knowledgeable Data Scientist.
Challenges	● Personas should be partially auto generated and presented to the Data Scientists prior to his analysis, based on certain similar aspects identified by the system (age group, location, interest, compensation requested, etc.).
GDPR Issue	● As regards the creation of Personas by DataVaults Data Scientist (SS3), it has to be further explored if this implies or not “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. ● Personas will be partially auto-generated based on certain similarity aspects identified by the system, such as age and compensation requested: it needs to be clarified if this implies or not an automated-decision making. It is important that, as already foreseen, the human intervention will be part of the task, especially in case some effects on the Individuals could occur (such as exclusion from some data sharing contract).
Ethical Issues	● As mentioned in the Scenarios driven by Individuals, it is relevant to consider also non-monetary rewarding mechanisms (SS4).
Privacy Issues	● The same considerations described in Scenario 10 apply driven by Data Seekers when they want to perform data analysis requests. ● User Privacy Exposure: The current user privacy risk exposure, as calculated by the DataVaults Risk Assessment framework, should not

	breach the privacy requirements (as defined by the user) when the Data Scientist creates the merged personas. This requires that quantified privacy risk exposure values are kept within the user acceptable boundaries; in any other case, DataVaults should inform the user of appropriate actions that should be taken for privacy enhancement. This also pertains to no malicious user or Observer gaining a significant advantage of identifying (de-anonymizing) the user by exploiting the trading mechanisms; linking number of transactions with (e.g.,) amount of (crypto) currency earned by an individual. ● User and Data Privacy: Privacy features, selected by the users, for sharing their data should be in line with the privacy preferences that were selected by the users when granting authorization to the Data Scientist for the creation of appropriate data personas. This entails that no advantage will be given to a Data Seeker or an Observer that have access to these personas, for being able to link data assets back to specific users (data owners).
Security Issues	● The same considerations described in Scenario 10 apply driven by Data Seekers when they want to data analysis requests.

4 FEATURES EXTRACTION

The Operations described in the DataVaults Methodology, as well as the steps of the high-level Operation Scenarios, imply a number of underlying functionalities that shall be provided by DataVaults. These functionalities, extracted from the combination of the methodology and the scenarios and formulated as features from the side of DataVaults, are presented in this section. Some features (for example, the "autopopulation of an Individual's profile") extend the core functionalities explicitly described in the methodology and scenarios and are listed as a possible addition to be considered for the DataVaults Platform. Finally, although user-management related tasks have not been included in the Methodology, are considered in this section as features, as they are a prerequisite for the overall functionality of the platform

The relation of the features to the methodology is denoted at Operation level, while the link to scenarios is also provided. A distinction has been made between features envisioned for the DataVaults Personal Application and those that shall be featured in the DataVaults Cloud Platform. This distinction is denoted in the name coding of each feature ("DVPERS_F_XX" and "DVPLAT_F_XX" respectively), as well as in the "Featured in" row. For features implying the involvement and communication of both sides, the "DVPLAT" indication has been used, as the Platform side will be delegated with the main workload for the operation.

DVPERS_F_01. Ability of Individual to register/login using third-party identity providers

Description	The Individual can register/login to the Personal DataVaults App using credentials provided through third party providers such as Facebook/Google/etc. and pre-fill part of his profile.
Featured In	Personal DataVaults App
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	Connection to third-party providers is available

DVPERS_F_02. Population of Individual's profile

Description	Upon the Individual's registration the DataVaults Personal App shall create the Individual's profile and populate it with the information s-/he inputs through the relevant input fields, including required basic information (as for example name, username, demographics etc.), but also some optional advanced fields related to hobbies and interests, preferences, education etc.
Featured In	Personal DataVaults App
Methodology Phase	N/A
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_03. AutoPopulation of Individual's profile

Description	The DataVaults Personal App shall extract profile information from the Individual's profile in linked data sources, such as Facebook, Fitbit and more, and enrich with this information the Individual's DataVaults personal profile.

Featured In	Personal DataVaults App
Methodology Phase	N/A
Related Scenarios	IND.01
Prerequisites	The Individual has linked sources to her/his profile (DVPERS_F_06, DVPERS_F_07)

DVPERS_F_04. Request for advanced profile information – Individual's profile enrichment

Description	The DataVaults Personal App from time to time shall prompt the Individual to input additional advanced information, through appropriate input forms such as ad-hoc questionnaires, that would help at a later stage in recommendations and the creation of Personas.
Featured In	Personal DataVaults App
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	N/A

DVPERS_F_05. Display profile completion status

Description	The DataVaults Personal App shall calculate and display to the Individual the level of completion of her/his profile.
Featured In	Personal DataVaults App
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	N/A

DVPERS_F_06. Authorise DataVaults Personal App to collect data from local data sources

Description	Upon registration the Individual is prompted to provide authorization to the DataVaults Personal App to access data sources located locally at the Individual's device and collect data, such as the "Images" folder. The Individual can authorize the DataVaults Personal App to access local data sources on her/his own initiative at any time.
Featured In	Personal DataVaults App
Methodology Phase	Phase I – Operation I.1
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_07. Connect Personal DataVaults App to external data sources

Description	To allow the retrieval of data from external data sources, the Individual is prompted to provide user credentials as tokens for the establishment of the API connection between the DataVaults Personal App and the external source.
Featured In	Personal DataVaults App
Methodology Phase	Phase I – Operation I.1
Related Scenarios	IND.01
Prerequisites	Connection to third-party providers is available

DVPERS_F_08. Data collection configuration options

Description	The DataVaults Personal App shall provide to the Individual a set of data collection configuration options regarding various aspects such as the data collection schedule and intervals, data retrieval access/user credentials expiration date, interval of data collection and more.
Featured In	Personal DataVaults App
Methodology Phase	Phase I – Operation I.1
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_09. Test data source connection and data availability

Description	Upon configuration of a new data retrieval process from a data source, the DataVaults Personal App shall make a test connection to verify the provided API connection details and retrieve a testing data sample to check data availability.
Featured In	Personal DataVaults App
Methodology Phase	Phase I – Operation I.2
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_10. Apply data quality check

Description	Each time a dataset is collected from a data source, the DataVaults Personal App shall perform the defined data quality check process to detect any inconsistencies, anomalies and errors in the retrieved data.
Featured In	Personal DataVaults App
Methodology Phase	Phase II – Operation II.1
Related Scenarios	IND.01
Prerequisites	A data quality check process has been defined by the DataVaults Data Scientist

DVPERS_F_11. Implement data transformations on fetched data

Description	The DataVaults Personal App shall perform the required data transformations to the collected data, based on the data quality check results. These data transformations include the application of the appropriate data cleaning and filtering techniques.
Featured In	Personal DataVaults App
Methodology Phase	Phase II – Operation II.1
Related Scenarios	IND.01
Prerequisites	Transformation workflows have been defined by the DataVaults Data Scientist

DVPERS_F_12. Mapping data to DataVaults schema

Description	The DataVaults Personal App shall perform the required mappings over the collected and pre-processed data to the DataVaults data schemas, based on

	mapping configurations already defined by the DataVaults Data Scientist, to allow queries and other operations.
Featured In	Personal DataVaults App
Methodology Phase	Phase II – Operation II.2
Related Scenarios	IND.01
Prerequisites	Mappings between the data schemas of the various sources to the DataVaults data model have been defined by the DataVaults Data Scientist

DVPERS_F_13. Semantic annotation/enrichment of data assets

Description	The DataVaults Personal App shall attach additional, machine-readable information to the retrieved data, regarding various concepts and domains, as well as annotate them with the appropriate labels, to semantically enrich them and facilitate other operations such as querying and more.
Featured In	Personal DataVaults App
Methodology Phase	Phase II – Operation II.2
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_14. Data assets linking

Description	The DataVaults Personal App shall create links between datasets acquired from different sources for the same individual, to enable the combination of data and the population of the unique Individual's data model instantiations, such as her/his personal profile or health data. Dataset linking will be done either via URIs or via the creation of multiple distributions of the same dataset.
Featured In	Personal DataVaults App
Methodology Phase	Phase II – Operation II.3
Related Scenarios	IND.01
Prerequisites	N/A

DVPERS_F_15. Local visualisation of datasets

Description	Certain types of predefined dataset types (coming from specific data sources) should be automatically visualized within the Personal DataVaults App to showcase to the user some information as extracted from these datasets, acting as a motivation point for the user to re-launch the App in the near future to witness his progress in those data categories. Such visualisations could include health/physical activity data, online behaviour data, etc.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.1
Related Scenarios	IND.02
Prerequisites	Data have been successfully collected from the linked sources

DVPERS_F_16. Running edge analytics

Description	The Personal DataVaults App should allow Individuals to choose some of the datasets locally stored and perform certain analytics tasks, based on pre-defined algorithms and few configuration inputs that would be available. The results of those could be visualized or presented as a file.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.1
Related Scenarios	IND.02
Prerequisites	Data have been successfully collected from the linked sources; Pre-defined data analytics tasks have been designed by the DataVaults Data Scientist

DVPERS_F_17. Generating assets from the outputs of edge analytics

Description	The Personal DataVaults App should allow users to save the outputs of their analyses as “assets” that can be stored/encrypted/shared/etc as any other dataset within the DataVaults ecosystem.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.1
Related Scenarios	IND.02
Prerequisites	Edge analytics have been executed

DVPERS_F_18. Encryption of personal data at the Individual’s side

Description	An encryption engine at the Individual’s side shall encrypt the personal data which are stored locally at the Individual’s side.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.2
Related Scenarios	IND.01
Prerequisites	Data have been successfully collected from the linked sources

DVPERS_F_19. Indexing of data assets at local storage

Description	Local secure storage facility at the side of the Individual shall incorporate a data catalogue service, which is used to document the type of data assets already stored in the local storage, providing an index of the actual data and other metadata information.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.4
Related Scenarios	IND.01, IND.02
Prerequisites	Data have been successfully collected from the linked sources; Encryption method that is applied on local storage, allows indexing

DVPERS_F_20. Storage of data assets at local storage

Description	Local secure storage facility shall persist the encrypted personal data assets, i.e. the retrieved personal data along with their metadata.
Featured In	Personal DataVaults App

Methodology Phase	Phase III – Operation III.3
Related Scenarios	IND.01, IND.02
Prerequisites	Data have been successfully collected from the linked sources

DVPERS_F_21. De-authorisation of connection with given data source

Description	Upon request from the Individual, the DataVaults Personal App will de-authorise the connection previously established with a given data source, thus stopping future retrieval of personal data from the given data source.
Featured In	Personal DataVaults App
Methodology Phase	Phase I – Operation I.3
Related Scenarios	IND.01
Prerequisites	A connection of the given data source and the DataVaults Personal App has been previously established

DVPERS_F_22. Removal of personal data from local storage

Description	Upon request from the Individual, the DataVaults Personal App will delete all data which have been previously collected from the given data source and is stored at the Individual's side.
Featured In	Personal DataVaults App
Methodology Phase	Phase III – Operation III.5
Related Scenarios	IND.01 IND.02
Prerequisites	Personal data have been previously collected by the DataVaults Personal App from the given data source.

DVPERS_F_23. Configuration of sharing anonymisation level

Description	The Individuals shall select the preferred level of anonymisation for the data asset they are going to share. They can choose to share it without applying anonymisation (eponymous), anonymise their data at an individual level (Digital Twin), or anonymise their data at a group level and make them available for the creation of Personas.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_24. Configuration of access policies

Description	The DataVaults Personal App shall provide the Individuals with a comprehensive access policy editor, to create manually the policies they would like to enforce to their shared data assets. The result will be a set of attribute-based rules that will be enforced by the Access Policy Engine upon every new data access request. Through the access policy editor, the Individual can also restrict all access, in case s-/he just wants to upload the

	data to the DataVaults Cloud Platform without making it available to any Data Seeker.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_25. Configuration of license(s)

Description	The Individual shall define the licensing terms under which s/he shares the data asset. The terms could affect aspects like using data for data analytics processes, sharing with other parties with/without notification of the owner, expiry of licence etc.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_26. Configuration of price tag(s)

Description	The Individual shall define the price that shall be paid by the Data Seeker to acquire the shared data asset under a specific license. For example, if the Data Seeker wants to buy a license allowing for a specific time period s/he would pay less than to buy a licence that doesn't expire.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_27. Suggestion of data asset price setting

Description	During the sharing configuration the DataVaults Platform shall assist the Individuals with the pricing scheme selection, by providing price suggestions based on other selected sharing aspects, such as the anonymisation and visibility levels.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.03
Prerequisites	N/A

DVPERS_F_28. Assessment of sharing privacy exposure

Description	The DataVaults Personal App shall provide the Individual with a Privacy Metrics Dashboard demonstrating the overall - current and projected - risk metrics based on the sharing configurations and the sensitivity of the enclosed information of both data already shared and data to-be-shared.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.2
Related Scenarios	IND.03

Prerequisites	N/A

DVPERS_F_29. Selection of DAA method at sharing by an Individual

Description	The DataVaults Personal App shall apply the direct anonymous attestation protocol to ensure the trustworthiness of the shared data as well as the Individual's device. The DataVaults Personal App shall provide the Individual with the option to bypass the direct attestation step, in order to speed up the sharing process. In this case the uploaded data assets will not be marked as "trusted".
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	The Individual's device supports the application of the DAA protocol

DVPERS_F_30. Save sharing configuration

Description	The DataVaults Personal App shall store the sharing configuration created by the Individual and make it available for future re-use as a sharing configuration template or/and modification of the sharing terms of this specific data asset.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_31. Use of existing sharing configuration template for a new data asset sharing job

Description	The Individuals shall be able to reuse saved sharing configurations from previously shared data assets. The users can select one of the saved configuration templates, review the various aspects and make any modifications, link it to the data asset they wish to share and finalise it to be executed at the time they have selected.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	Configuration templates have been saved during past data asset sharing configurations

DVPERS_F_32. Edits on saved sharing configuration

Description	The DataVaults Personal App shall allow the Individual to edit the various parameters of the stored sharing configuration. These modifications will be put into force for any new data requests but will not affect already valid contracts that were created with the previous sharing configuration.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04

Prerequisites	Configuration templates have been saved during past data asset sharing configurations

DVPERS_F_33. Generation and management of a sharing schedule

Description	The DataVaults Personal App shall provide a sharing scheduling functionality, to enable the Individual to automatically share her/his data assets, under the defined sharing terms.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	N/A

DVPERS_F_34. Management of private DataVaults contract for data asset transfer

Description	Whenever a data asset is uploaded from the DataVaults Personal App to the DataVaults Cloud Platform, a contract is issued between the two parties, encompassing the sharing configurations made by the Individual. The contract remains valid until the set expiration date -if any applies – or until the Individual (with a direct action (change of sharing configuration) or indirectly through the propagation of another action (such as for example the right to be forgotten) changes the sharing terms under effect. In the second case the previous DataVaults contract is invalidated and a new one is appended on top of it that includes the latest changes.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.3
Related Scenarios	IND.05
Prerequisites	A data asset sharing process has been configured

DVPERS_F_35. Individuals transaction log

Description	The Individual can view at any time the transactions he has taken part in and witness the actual value s-/he has gained from the shared data assets that have been purchased by Data Seekers.
Featured In	Personal DataVaults App
Methodology Phase	Phase VIII – Operation VIII.2
Related Scenarios	IND.02
Prerequisites	N/A

DVPERS_F_36. Information on the Digital Twin Identities owned by the Individual

Description	The Individual can view at any time under which Digital Twin Identities he has shared data anonymously with the DataVaults Cloud Platform.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.02
Prerequisites	N/A

DVPERS_F_37. Notifications generated by DataVaults Personal App

Description	DataVaults Personal App will create notifications based on the activity of the Individual, which will be dispatched to the Individual via the DataVaults Personal App. E.g. ● Successful connection with data source ● De-authorisation of connection with data source ● Removal of personal data assets from local storage ● Notification for updated risk exposure metrics ● Private contract successful establishment ● Successful sharing with DataVaults Cloud Platform ● Revocation of data asset sharing from DataVaults Cloud Platform ● Notification for a Data Request
Featured In	Personal DataVaults App
Methodology Phase	Phase VIII – Operation VIII.1
Related Scenarios	Horizontal
Prerequisites	N/A

DVPLAT_F_01. Transferring value of shared data assets to the Individual

Description	The DataVaults Platform shall provide a distributed ledger-enabled compensation mechanism that will be activated whenever a shared asset is bought by a Data Seeker. This mechanism encompasses the transfer of the agreed amount of currency from the DataVaults Cloud Platform to the Individual's Personal DataVaults Wallet and the issuance of a contract between the two parties, that will record the transaction and the agreed terms.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase IV – Operation IV.3
Related Scenarios	DSE.02, DSE.03
Prerequisites	Currency has been transferred from the Data Seeker to the DataVaults platform

DVPLAT_F_02. Generation of Digital Twin

Description	The DataVaults Cloud Platform shall create the Digital Twin of an Individual, whenever this is delimited by the data asset sharing configuration. The Digital Twin is generated by anonymizing and obfuscating personally identifiable data while preserving the valuable information enclosed in the data asset, using the identity provided by the Identities Wallet.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase IV – Operation IV.1
Related Scenarios	IND.04
Prerequisites	The Individual has selected during the sharing configuration to share data as a Digital Twin

DVPLAT_F_03. Generation of Personas

Description	The DataVaults Cloud Platform shall provide the DataVaults Data Scientist with an engine for the creation of aggregated profiles composed of data assets from different Individuals that share certain similarities, the so-called Personas. The Individuals have indicated in the sharing configuration their intention to share data for this purpose. The privacy of the Individuals 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.
Featured In	Personal DataVaults App
Methodology Phase	Phase IV – Operation IV.3, Phase VII – Operation VII.3
Related Scenarios	DVO.01
Prerequisites	The Individual has selected during the sharing configuration to share data as a Persona

DVPLAT_F_04. Enforcement of data access policies

Description	The DataVaults Cloud Platform shall facilitate the enforcement of access policies upon every data access request. These access policies have been created by the Individual through the Access Policy Editor during the sharing configuration. The access policy engine is updated with the respective policies when the data assets are uploaded to the DataVaults Cloud platform. Every time an access request is made, these access policies are resolved to decide on granting or denying access to the requester.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase V – Operation V.1
Related Scenarios	IND.05
Prerequisites	The Individual has defined during the sharing configuration any data access policies that shall apply

DVPLAT_F_05. Encryption of data assets at the DataVaults Cloud Platform

Description	An encryption engine at the DataVaults Cloud Platform shall: ● accommodate encrypted data assets coming from the Individual's side ● encrypt personal data which are shared with and stored at the DataVaults Cloud Platform ● build an Encrypted Searchable Data Lake using encrypted data assets as described above
Methodology Phase	Phase V – Operation V.1
Related Scenarios	IND.05
Prerequisites	N/A

DVPLAT_F_06. Enable searching over shared encrypted data assets

Description	The DataVaults Cloud Platform shall employ techniques which allow Data Seekers to search through data assets shared with the DataVaults Cloud

	Platform, even when those data assets are encrypted and therefore their content is not disclosed to the Data Seeker.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase V – Operation V.2, Operation V.3
Related Scenarios	IND.05
Prerequisites	Implement encryption techniques such as Searchable Encryption

DVPLAT_F_07. Storage of data assets at DataVaults Cloud Platform

Description	Secure storage facility of the DataVaults Cloud Platform shall persist the encrypted personal data assets, i.e. the retrieved personal data along with their metadata.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase V – Operation V.1
Related Scenarios	IND.05
Prerequisites	Data assets have been shared with the DataVaults Cloud Platform by the given Individual.

DVPLAT_F_08. Indexing of data assets at DataVaults Cloud Platform

Description	DataVaults Cloud Platform storage shall incorporate a data catalogue service, which is used for the application of data indexing techniques over the data assets and their metadata shared with the DataVaults Cloud Platform. Different types of indexes can be used for this purpose, depending on the applicable data model and metadata.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase V – Operation V.3
Related Scenarios	IND.05
Prerequisites	N/A

DVPLAT_F_09. Removal of data assets from DataVaults Cloud Platform

Description	Upon request from the Individual, the DataVaults Cloud Platform will delete all data which have been previously shared from the given Individual and is stored at DataVaults Cloud Platform, apart from the ones which are already part of active public contracts.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase V – Operation V.4
Related Scenarios	IND.06
Prerequisites	Data assets have been previously shared with the DataVaults Cloud Platform by the given Individual.

DVPLAT_F_10. Data Seekers profile editing

Description	Upon registration the Data Seekers are prompted to provide some required basic information regarding the affiliated organization, such as the organisation's full name, type, description of the organisation's scope and more. The organisation's profile can be enriched with additional, optional information by the Data Seeker at a later stage.

Featured In	DataVaults Cloud Platform
Methodology Phase	N/A
Related Scenarios	DSE.01
Prerequisites	N/A

DVPLAT_F_11. Data Seekers profile authenticity validation

Description	The DataVaults Cloud Platform shall validate the Data Seeker is truly who he/she claims to be, through a validation process requiring the submission of relevant documents and the manual validation by an operator from the side of the DataVaults Cloud Platform.
Featured In	DataVaults Cloud Platform
Methodology Phase	N/A
Related Scenarios	DSE.01
Prerequisites	N/A

DVPLAT_F_12. Apply DAA method to Data Seeker connection

Description	The DataVaults Cloud Platform shall verify the trustworthiness and validity of the Data Seeker's device prior to allowing them join the DataVaults framework, through the Direct Anonymous Attestation functionality of TPMs.
Featured In	DataVaults Cloud Platform
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	The Data Seeker's device supports the application of the DAA protocol

DVPLAT_F_13. Construction of simple or complex data asset queries

Description	The DataVaults Cloud Platform shall facilitate the Data Seekers in searching for data assets of their interest. The Data Seekers can create through the available query editor and forms, queries serving various types of searches (name search, range queries, metadata-based search, search based on sharing settings etc.).
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.1
Related Scenarios	DSE.01
Prerequisites	N/A

DVPLAT_F_14. Navigation through query results in the unencrypted data lake

Description	The DataVaults Cloud Platform shall find data assets matching or approximating the query parameters with a satisfying proximity score, always in collaboration with the access policy engine and return the search results to the Data Seeker. The Data Seeker can navigate through the query response using various views (such as lists and filters) and explore the returned data assets (if s-/he has already acquired them) or the previews (for data assets not yet fully available to her/him).
Featured In	DataVaults Cloud Platform

Methodology Phase	Phase VI– Operation VI.1
Related Scenarios	DSE.01
Prerequisites	N/A

DVPLAT_F_15. Navigation through query results in the encrypted data lake

Description	The DataVaults Cloud Platform shall help the Data Seekers to locate possible data assets of their interest in the encrypted data lake. It employs a searchable encryption mechanism to facilitate searches over encrypted data, and in collaboration with the access policies engine returns to the Data Seekers the metadata and data samples of the located data assets. The Data Seeker can navigate using various views and filters through the response to her/his request.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI– Operation VI.1
Related Scenarios	DSE.01
Prerequisites	Data assets in the encrypted data lake are searchable

DVPLAT_F_16. Custom data asset request from a Data Seeker

Description	The DataVaults Cloud Platform shall enable the Data Seeker to request data assets not yet (fully) shared by an Individual. For this purpose, the Data Seeker shall create an offer that incorporates the various sharing aspects (price, license terms etc.). This request will be transformed by the DataVaults Cloud platform to a contract that is afterwards reviewed by the Individual who can either accept the terms, share the data asset and validated it or reject it.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.1, Phase IV– Operation IV.3
Related Scenarios	DSE.03
Prerequisites	Data Seeker can view previews or partial information that does not hinder the privacy of the Individual

DVPLAT_F_17. Transferring value for acquired data assets to the DataVaults Platform

Description	The DataVaults Platform shall provide a distributed ledger-enabled mechanism that will be activated whenever a shared asset is bought by a Data Seeker. This mechanism encompasses the transfer of the agreed amount of currency from the Data Seeker to the DataVaults Cloud Platform.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase IV – Operation IV.4
Related Scenarios	DSE.02, DSE.03
Prerequisites	N/A

DVPLAT_F_18. Management of public DataVaults contract for data asset acquisition

Description	The DataVaults Cloud Platform shall issue and validate a public ledger contract between itself and the Data Seeker, whenever s-/he acquires a shared data asset. The contract is validated once the data asset and

	cryptocurrency transactions are completed and remains valid for the time period defined by the terms.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI– Operation VI.2
Related Scenarios	DSE.02, DSE.03
Prerequisites	N/A

DVPLAT_F_19. Data Seekers transaction log

Description	The DataVaults Cloud Platform shall provide the Data Seekers with a log of the performed currency transactions (from the Data Seeker to the DataVaults Cloud Platform), along with the transaction details (amount, date of transaction, related data asset etc.). The Data Seekers can select from the provided filtering and sorting options for improved navigation.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.1, Phase VI – Operation VI.2
Related Scenarios	N/A
Prerequisites	N/A

DVPLAT_F_20. Data Seekers assets vault

Description	The DataVaults Cloud Platform shall enable the Data Seekers to explore all data assets they have acquired or have uploaded, in a dedicated Data Vault. The Data Seekers can view the complete data asset list or apply filters and sorting options for their convenience.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_21. Present disk usage quotas to Data Seekers

Description	The DataVaults Cloud Platform shall inform the Data Seekers of the available cloud space they are entitled to.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.3
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_22. Upload own data to the Data Seekers assets vault

Description	The DataVaults Cloud Platform shall enable the Data Seekers to upload data from their external data repositories to their Data Vault. These data can be used in data analytics, combined with data assets acquired through DataVaults.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.3, Phase VI – Operation VI.3
Related Scenarios	DSE.04
Prerequisites	Connection to Data Seeker’s external data sources is available

DVPLAT_F_23. Exporting acquired data assets from the DataVaults Cloud Platform

Description	The DataVaults Cloud Platform shall provide the Data Seekers with the option to download and store outside the DataVaults environment any data assets they are eligible for, as a file. The export mechanism shall allow for various export formats to choose from, always respecting the licensing terms.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.3
Related Scenarios	DSE.02, DSE.03
Prerequisites	Data shared through the DataVaults Cloud have been acquired by the Data Seeker

DVPLAT_F_24. Remote access to acquired data assets via the DataVaults Cloud Platform

Description	The DataVaults Cloud Platform shall enable the Data Seekers to remotely access data assets they have in their Data Vault, through available APIs. The Data Seekers shall configure the required API details to enable DataVaults connect to their application.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VI – Operation VI.3
Related Scenarios	DSE.01, DSE.04
Prerequisites	Connection to the DataVaults Cloud has been established

DVPLAT_F_25. Present analytics time quotas to Data Seekers

Description	The DataVaults Cloud Platform shall inform the Data Seekers of the available processing time they are entitled to, to perform their analytics.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII – Operation VII.1
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_26. Configuration of data analytics task

Description	The DataVaults Personal App/Cloud platform shall allow the user to create and configure an analytics task using an algorithm selected from a list of supported operations to run over a specific data asset.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_27. Creation of a chain of data analytics tasks

Description	The DataVaults Cloud platform shall allow the user to create a chain of successive data analytics operations to be performed on a data asset. Each operation will have as input the results of the previous operation in the chain.

Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_28. Compatibility check of data assets with selected analytics task(s)

Description	The DataVaults Cloud platform shall review the configured data analytics task and check whether it can be performed on the specific data type and format. This check also happened for analytics chains, where DataVaults compares the expected inputs and outputs of the successive algorithms and indicates any possible incompatibilities. If required and applicable, certain transformations are applied over the data prior to being fed to the analytics playground.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured by the Data Seeker

DVPLAT_F_29. Storage of data analytics task configuration

Description	The DataVaults Cloud platform shall enable the user to store the analytics configuration, so that it is available for future use.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured by the Data Seeker

DVPLAT_F_30. Suggestion of pre-defined data analytics chains

Description	The DataVaults Personal App/Cloud platform shall help the user in defining an algorithmic chain that would makes sense, by providing a number of pre-defined data analytics tasks combinations to choose from.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04, DVO.05
Prerequisites	Pre-defined analytics tasks have been designed by the DataVaults Data Scientist

DVPLAT_F_31. Visualisation of data analysis results on a dashboard

Description	The DataVaults Cloud platform shall provide the user with a visualization dashboard supporting various types of graphic representations, to facilitate better understanding of the data and the analysis.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.2
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured and executed

DVPLAT_F_32. Customisation of visualisation dashboard

Description	The DataVaults Cloud platform shall enable the user to have a visualization dashboard that better fits her/his needs. The user can customize the types of visualisations available in the main dashboard through a friendly user interface. This customisable dashboard can be modified at any time by the user.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.2
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_33. Storage of visualisation configuration

Description	The DataVaults Cloud platform shall enable the user to store the visualisation configuration, so that it is available for future use.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.2
Related Scenarios	DSE.04
Prerequisites	A visualisation configuration has been defined by the Data Seeker

DVPLAT_F_34. Storage of data analysis results

Description	The DataVaults Cloud platform shall make the analytics results available to the user for future reference, by providing the option to store them as a project in DataVaults.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured and executed

DVPLAT_F_35. Storage of visualizations as projects

Description	The DataVaults Cloud platform shall make the visualisation results available to the user for future reference, by providing the option to store them as a project in DataVaults.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.2
Related Scenarios	DSE.04
Prerequisites	The visualisation of an analytics task has been configured and executed

DVPLAT_F_36. Availability of data analysis results for download in various formats

Description	The DataVaults Cloud platform shall enable the user to store the results of the performed analysis locally. The user shall select the preferred export format (ex. machine-readable formats, such as JSON, xml, or other file types such as PDF).
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.3
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured and executed

DVPLAT_F_37. Availability of data analysis results through an API

Description	The DataVaults Cloud platform shall enable the user to acquire the results of the performed analysis through the available APIs.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.3
Related Scenarios	DSE.04
Prerequisites	An analytics task has been configured and executed; Connection to the Data Seeker's APIs has been established

DVPLAT_F_38. Availability of visualization results for download in various formats

Description	The DataVaults Cloud platform shall enable the user to store the results of the performed visualization, locally. The user shall select the preferred export format (ex. image, PDF).
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.3
Related Scenarios	DSE.04
Prerequisites	The visualisation of an analytics task has been configured and executed

DVPLAT_F_39. Complete erasure of data from the Analytics Playground

Description	The DataVaults Personal App/Cloud platform shall provide the option to discard any traces of data from the secure analytics playground, in case the user wishes so. This pertains both the input of the data analytics task as well as the analytics results and visualisations.
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VII– Operation VII.1, Phase VII– Operation VII.2
Related Scenarios	DSE.04
Prerequisites	N/A

DVPLAT_F_40. Notifications generated by DataVaults Cloud Platform

Description	DataVaults Cloud Platform will create notifications based on the activity of the Data Seeker and/or the Individual, which will be dispatched to the Data Seeker via the DataVaults Cloud Platform or to the Individual via the DataVaults Personal App. E.g. ● Success of a requested data analysis job ● Public contract successful establishment ● Public contract expiration notification ● Private contract updating ● Success/Failure of data assets acquisition ● Acceptance/Rejection of data asset acquisition proposal
Featured In	DataVaults Cloud Platform
Methodology Phase	Phase VIII – Operation VIII.1

Related Scenarios	Horizontal
Prerequisites	N/A

DVPLAT_F_41. Right to be forgotten

Description	The user can at any time choose to delete all her/his data from the DataVaults ecosystem and leave the platform, while the platform will only retain the public contracts that are active.
Featured In	DataVaults Cloud Platform
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	User profile data can be removed; Retrieved and uploaded data assets can be completely removed from the DataVaults Platform

DVPLAT_F_42. Right to data portability

Description	The user can at any time choose to obtain and reuse all her/his data from the DataVaults ecosystem and move/copy/transfer them from DataVaults to another IT environment by downloading/exporting them or through an API connection.
Featured In	DataVaults Cloud Platform
Methodology Phase	N/A
Related Scenarios	N/A
Prerequisites	Data in the DataVaults Personal App and in the DataVaults Wallet can be exported in various forms

5 THE DATAVAULTS MVP – VERSION #1

5.1 WHAT IS AN MVP

A Most Valuable Product (MVP) can be seen both as a product and as a development technique that aims to deliver the maximum value to the involved stakeholders (demonstrators and technical partners) both during and post project implementation, given any constraints placed upon it. Such constraints refer to the available time and budget, technical constraints and the project's overall scope. As such, the Most Valuable Product can be considered as an alternate to the more common Minimum Viable Product, an approach towards product design and planning that is coined to Frank Robinson. Minimum Viable Product is typically a version of a product with just enough features to satisfy early adopters and provide feedback for future product development. It can be used for gathering insights early during product development; doing so is often less expensive than developing a product with more features and wait to receive feedback at the end of product development. However, this approach could mislead the development teams to stripping down the released product from any features that are considered non-basic, in order to reduce time-to-market, thus leading to entering the market early but lacking true value and differentiation from the competitors. Additionally, the mentality of treating early adopters as testers instead of users can be inherently problematic, as it moves the accountability for delivering a mature product from the development team to the users, and could lead to not being able to retain them by offering only basic functionalities with the promise to improve in the future.

Considering the above, DataVaults has instead adopted the approach of Most Valuable Product development, that in its core is oriented at validating the envisioned solution, instead of identifying problems. In essence, the DataVaults MVP represents the overall mindset and process adopted for product development to consider user expectations, deliver actual value and validate the methodological ideas and hypothesis. The DataVaults MVP is expected to become instrumental in guiding the design and development activities throughout the project implementation and represents the basis of the platform release that will be delivered.

The process towards developing the DataVaults MVP (Figure 3) depends heavily on having a clear definition of the scope and purpose of the project, by fleshing out the actual value of the proposed solution for the end users. The demonstrators play a crucial role in this process. Initially, they provided their insights from the domain, as they described in detail current processes and any challenges they face. This facilitated the identification of existing space for procedural redesign in the Methodology and scenarios, to showcase how the common workflows can become more efficient with the intervention of DataVaults. Furthermore, by recognizing existing gaps, DataVaults has also designed added value services to provide end-user organisations with competitive advantages. The next step towards the DataVaults MVP is the assessment of value of the DataVaults functionalities that have been derived from the DataVaults Methodology and the respective scenarios as a set of homogenised features. The expected business value of these features is assessed by the demonstrators through a voting procedure. The technical partners on the other hand will also vote on the value and complexity of the features from the view of implementation, thus highlighting any technical prerequisites

and constraints. The combination of these two, leads to the consolidated DataVaults MVP that will drive the technical requirements extraction, the architectural design, the actual development, and the integration of the DataVaults solution in the next phases of the project. In this context, it needs to be noted that even if the MVP pinpoints some features as delivering the most value or being necessary for DataVaults to be deployed and validated, it does not dictate the DataVaults consortium to seize their work when implementing them; on the contrary, the MVP is a strategy for distributing wisely the development and integration workload, focusing on the most valuable assets, while taking into account complexity constraints that could push some features for later stages of the project in order to provide time for research or reach a satisfying level of maturity for the other features.

The present section of deliverable D1.3 focuses in the second and the third step towards the MVP, namely at assessing the added value of the set of features identified in section 4 and perform a preliminary prioritization of those. This assessment is conducted within the consortium for version I of the platform, with the plan to expand to external stakeholders for version II of the platform. During MVP definition it will be taken into consideration that the support of certain activities/features is mandatory for the implementation of others which are dependent on them and is expected to evolve into prioritised user stories in WP5 and in particular in deliverable D5.1. In summary, as depicted in the following figure, the approach that is applied for the MVP definition bears three core phases, namely Feature Definition, Feature Assessment and MVP Consolidation.

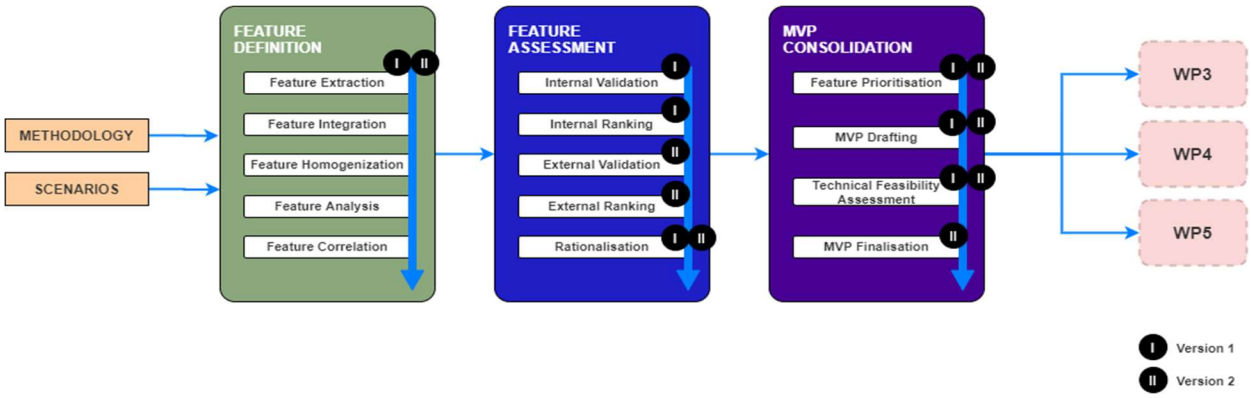


Figure 3 - DataVaults MVP Approach

5.2 Features Value Assessment

The assessment of the extracted features happens from a business and a technical view. Both demonstrators and technical partners are involved to provide their domain expertise and evaluate independently the individual features, using two popular prioritisation methods. The results of the two voting procedures are correlated in the end, to end up with the final ranking of the features.

The business value assessment of the features is performed by the demonstrators using the MoSCoW method. The name of this method stands for the four labels that are used to categorise the features, namely: Must-have, Should-have, Could-have and Won't have. Although this is a method usually employed from development teams to label the importance

of user stories, in our case it has been adapted to reflect the anticipated business value added by the feature for the demonstrators. More specifically, the labels refer to:

- **Must-have**: A feature that is of critical value for the demonstrator and needs to be implemented for the project to be successful. Such features should be limited to only a minimal set, in order for the project to be viable.
- **Should-have**: Should-have features add important value to the demonstrator's procedures and need to be implemented for the project to succeed but are usually less time-critical and can be implemented at a later stage.
- **Could-have**: Could-have features improve the overall user experience but are not critical for the demonstrator. They are usually simple upgrades and nice-to-have extras, like improving error messages or adjusting what is viewed on a screen on certain devices.
- **Won't have**: These are features with the lowest business value for the demonstrator and usually are left for the last stages of a project or remain at the backlog due to a low effort-to-value ratio.

The five DataVaults demonstrators were prompted to add a MoSCoW label to each feature along with any comments to justify their rating. Once this step was completed, it was followed by the processing of the ratings for the extraction of one label per feature. The four labels of the MoSCoW model were matched to a scale from one to four, with 'Won't have' being equal to one, 'Could have' scoring two and so on. As the vote of each demonstrator is of equal weight, the consolidation of their ratings was performed with a simple computation of the average for each feature. For the reverse mapping of the computation's result to the MoSCoW scale another approach was followed. Taking into account the definition stating that 'Must-have' features should be limited to the truly crucial ones, only those that were unanimously rated as such by all five demonstrators, thus scoring an average equal to four, were labelled as a 'Must-have'. The rest of the features were rated as follows:

Table 2 - Reverse Mapping of Average Feature Scores to MoSCoW Labels

Average Score	Final Label
<2	Won't-have
<3	Could-have
<4	Should-have
=4	Must-have

From the final results of the demonstrators' voting (visually presented in Figure 4) it is interesting to mention that no feature scored under two, thus having no features labelled as 'Won't-have'. This could be explained by the level of granularity of the described features.

On the contrary, the extracted features stay at a higher level, and describe functionalities deriving from the scenarios and methodology, thus not being easy to be left aside as 'redundant'. As expected, only five out of 79 features where unanimously voted as 'Must-have'. The table of features along with the final MoSCoW labels will be included in the next section presenting the MVP Consolidation.

For the second part of the assessment, a complexity matrix is employed, comprising two binary factors -value and complexity. Features are categorized in four types, as resulting from the four possible combinations of rankings, namely

- 'highly valuable and complex',
- 'highly valuable and not complex',
- 'not highly valuable and complex' and
- 'not highly valuable and not complex'.

In this context, value is not solely related to added business value, but is also interpreted in the light of added technical value, as for example some features that seem unimportant from a business scope, yet might be crucial as the basis or a prerequisite for the development of the most prominent functionalities of a product. This criterion plays an important role in defining the 'urgency' of a specific feature, in relation to the others in the backlog. Complexity is another factor that determines the planning in terms of time and resources that will be dedicated to the development of a functionality. Together, these two factors aim to spot which features are crucial to be handled from the early steps of the development, which can be easily implemented to enrich the first release and outline those that may need more research and development effort but have a high value-to-effort rate.

The twelve technical partners of DataVaults participated in this part of the prioritization process. Although each of them is responsible for the implementation of specific components, they were asked to vote for all the features, relevant or not to their individual work in DataVaults platform, as their experience allows them to have a good overall perception of the anticipated value and complexity. Again, the vote of each partner contributes equally to the final results. The following process was applied to extract the combined results: the values of the two voting criteria were mapped to 0 or 1 (highly valuable = 1, not highly valuable = 0, complex = 1, not complex = 0), and then the average value of the twelve votes was computed for each feature.

The following figure demonstrates that final results from the two voting rounds. The features are placed on a plane defined by the two axes (value and complexity), based on their pair value, to gain a better visual overview of their distribution. Colour coding is used to easily distinguish the four areas of the plane. Every feature is also tagged with a colour indicating its MoSCoW label (green for 'Must-haves', yellow for 'Should-haves' and red for 'Could-haves').

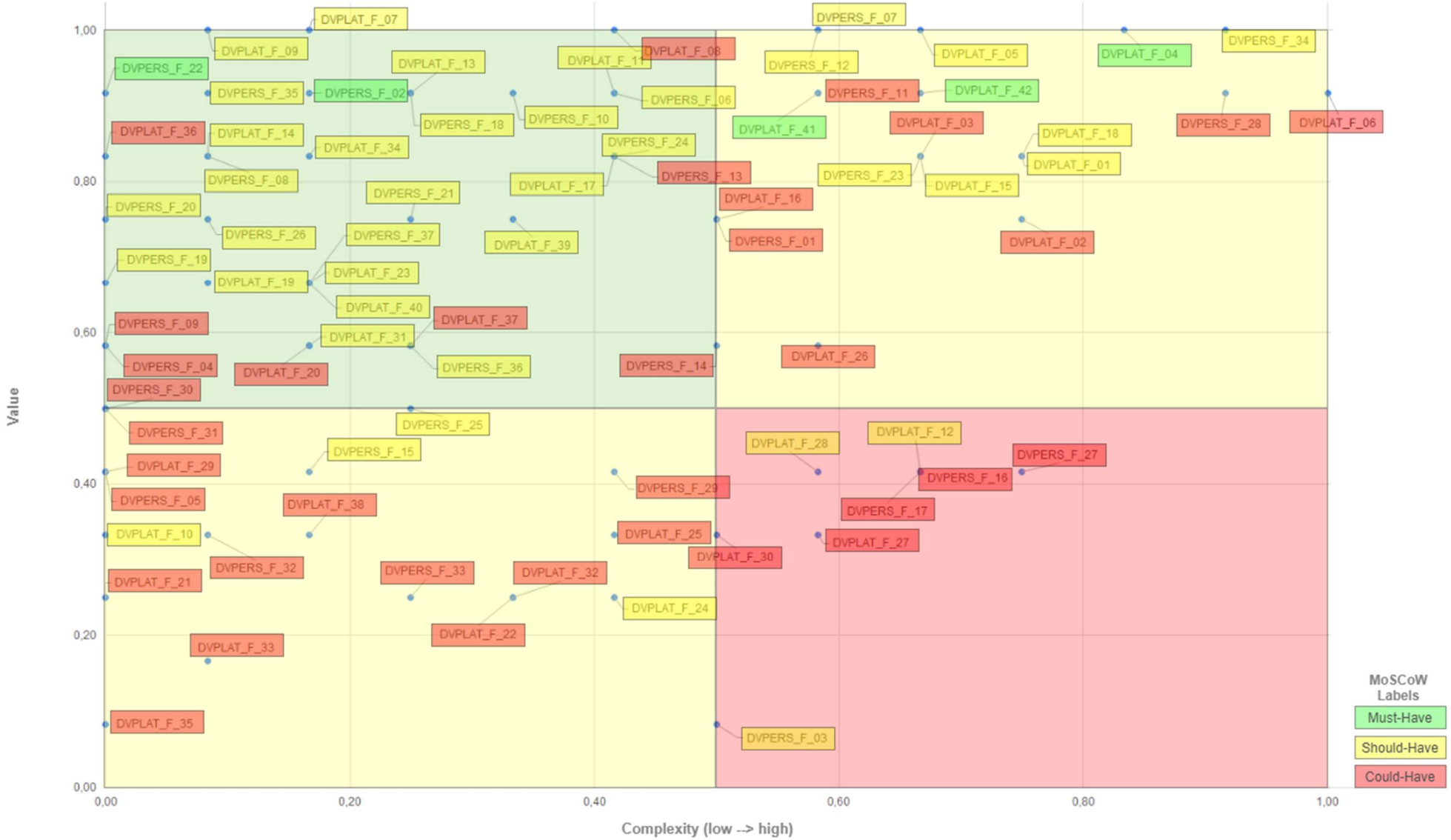


Figure 4 - Visualisation of Feature Voting Results

5.3 MVP CONSOLIDATION – VERSION #1

The consolidated DataVaults MVP consists of the extracted features, grouped in the four categories as resulting from the technical voting, and tagged with the MoSCoW label that was assigned to them by the pilots. The features in each Table are sorted in ascending technical value, to provide an initial prioritisation. It is worth mentioning, that no features from the initial list have been excluded from the MVP, as they were all rated as 'Could-haves' or above from the pilots. The main focus of development will be placed on 'Must-have' and 'Should-have' features and features that were rated as highly valuable by the technical partners, to ensure the delivery of a complete and highly valuable final MVP. The technical complexity of some highly valuable features might affect their planning towards the second release of DataVaults, as they require more effort and research. Furthermore, 'Could-have' features that also scored as 'not highly valuable', will not be left out, as they will enrich the final product as nice-to-have extras for the end users.

The following tables constitute the first version of the DataVaults MVP. These will be revised in the context of deliverable D1.4 to reflect the results of the second feature value assessment that will also involve external actors. Additionally, any additions or updates to features that will occur based on the implementation, integration and deployment of the first version of the DataVaults platform, will be included in the second version of the MVP.

Table 3 - Highly Valuable and Not Complex

Feature ID	Feature Title	Pilot's View
DVPERS_F_30	Save sharing configuration	Could-have
DVPERS_F_31	Use of existing sharing configuration template for a new data asset sharing job	Could-have
DVPERS_F_25	Configuration of license(s)	Should-have
DVPERS_F_04	Request for advanced profile information – Individual's profile enrichment	Could-have
DVPERS_F_09	Test data source connection and data availability	Could-have
DVPLAT_F_20	Data Seekers assets vault	Could-have
DVPLAT_F_31	Visualisation of data analysis results on a dashboard	Should-have
DVPERS_F_36	Information on the Digital Twin Identities owned by the Individual	Should-have
DVPLAT_F_37	Availability of data analysis results through an API	Could-have
DVPERS_F_19	Indexing of data assets at local storage	Should-have
DVPLAT_F_19	Data Seekers transaction log	Should-have
DVPERS_F_37	Notifications generated by DataVaults Personal App	Should-have
DVPLAT_F_23	Exporting acquired data assets from the DataVaults Cloud Platform	Should-have
DVPLAT_F_40	Notifications generated by DataVaults Cloud Platform	Should-have
DVPERS_F_20	Storage of data assets at local storage	Should-have
DVPERS_F_26	Configuration of price tag(s)	Should-have
DVPERS_F_21	De-authorisation of connection with given data source	Should-have
DVPLAT_F_39	Complete erasure of data from the Analytics Playground	Should-have
DVPLAT_F_36	Availability of data analysis results for download in various formats	Could-have
DVPERS_F_08	Data collection configuration options	Should-have

DVPLAT_F_14	Navigation through query results in the unencrypted data lake	Should-have
DVPLAT_F_34	Storage of data analysis results	Should-have
DVPERS_F_13	Semantic annotation/enrichment of data assets	Could-have
DVPERS_F_24	Configuration of access policies	Should-have
DVPLAT_F_17	Transferring value for acquired data assets to the DataVaults Platform	Should-have
DVPERS_F_22	Removal of personal data from local storage	Must-have
DVPERS_F_35	Individuals transaction log	Should-have
DVPERS_F_02	Population of Individual's profile	Must-have
DVPERS_F_18	Encryption of personal data at the Individual's side	Should-have
DVPLAT_F_13	Construction of simple or complex data asset queries	Should-have
DVPERS_F_10	Apply data quality check	Should-have
DVPERS_F_06	Authorise DataVaults Personal App to collect data from local data sources	Should-have
DVPLAT_F_11	Data Seekers profile authenticity validation	Should-have
DVPLAT_F_09	Removal of data assets from DataVaults Cloud Platform	Should-have
DVPLAT_F_07	Storage of data assets at DataVaults Cloud Platform	Should-have
DVPLAT_F_08	Indexing of data assets at DataVaults Cloud Platform	Could-have

Table 4 - Highly Valuable and Complex

Feature ID	Feature Title	Pilot's View
DVPERS_F_14	Data assets linking	Could-have
DVPLAT_F_26	Configuration of data analytics task	Could-have
DVPERS_F_01	Ability of Individual to register/login using third-party identity providers	Could-have
DVPLAT_F_16	Custom data asset request from a Data Seeker	Could-have
DVPLAT_F_02	Generation of Digital Twin	Could-have
DVPERS_F_23	Configuration of sharing anonymisation level	Should-have
DVPLAT_F_03	Generation of Personas	Could-have
DVPLAT_F_15	Navigation through query results in the encrypted data lake	Should-have
DVPLAT_F_01	Transferring value of shared data assets to the Individual	Should-have
DVPLAT_F_18	Management of public DataVaults contract for data asset acquisition	Should-have
DVPERS_F_11	Implement data transformations on fetched data	Could-have
DVPLAT_F_41	Right to be forgotten	Must-have
DVPLAT_F_42	Right to data portability	Must-have
DVPERS_F_28	Assessment of sharing privacy exposure	Could-have
DVPLAT_F_06	Enable searching over shared encrypted data assets	Could-have
DVPERS_F_07	Connect Personal DataVaults App to external data sources	Should-have
DVPERS_F_12	Mapping data to DataVaults schema	Should-have
DVPLAT_F_05	Encryption of data assets at the DataVaults Cloud Platform	Should-have
DVPLAT_F_04	Enforcement of data access policies	Must-have
DVPERS_F_34	Management of private DataVaults contract for data asset transfer	Should-have

Table 5 - Not Highly Valuable and Not Complex

Feature ID	Feature Title	Pilot's View
DVPLAT_F_35	Storage of visualizations as projects	Could-have
DVPLAT_F_33	Storage of visualisation configuration	Could-have
DVPLAT_F_21	Present disk usage quotas to Data Seekers	Could-have
DVPERS_F_33	Generation and management of a sharing schedule	Could-have
DVPLAT_F_22	Upload own data to the Data Seekers assets vault	Could-have
DVPLAT_F_32	Customisation of visualisation dashboard	Could-have
DVPLAT_F_24	Remote access to acquired data assets via the DataVaults Cloud Platform	Should-have
DVPLAT_F_10	Data Seekers profile editing	Should-have
DVPERS_F_32	Edits on saved sharing configuration	Could-have
DVPLAT_F_38	Availability of visualization results for download in various formats	Could-have
DVPLAT_F_25	Present analytics time quotas to Data Seekers	Could-have
DVPERS_F_05	Display profile completion status	Could-have
DVPLAT_F_29	Storage of data analytics task configuration	Could-have
DVPERS_F_15	Local visualisation of datasets	Should-have
DVPERS_F_29	Selection of DAA method at sharing by an Individual	Could-have

Table 6 - Not Highly Valuable and Complex

Feature ID	Feature Title	Pilot's View
DVPERS_F_03	AutoPopulation of Individual's profile	Should-have
DVPLAT_F_30	Suggestion of pre-defined data analytics chains	Could-have
DVPLAT_F_27	Creation of a chain of data analytics tasks	Could-have
DVPLAT_F_28	Compatibility check of data assets with selected analytics task(s)	Should-have
DVPERS_F_16	Running edge analytics	Could-have
DVPERS_F_17	Generating assets from the outputs of edge analytics	Could-have
DVPLAT_F_12	Apply DAA method to Data Seeker connection	Should-have
DVPERS_F_27	Suggestion of data asset price setting	Could-have

6 CONCLUSIONS AND NEXT STEPS

Deliverable D1.3 documented the work and outcomes of T1.5. These include the definition of the DataVaults integrated methodology, the drafting of the DataVaults high level operation scenarios and the consolidation of the DataVaults Most Valuable product. The outcomes of this deliverable were produced sequentially, based on the following approach:

Initially, the DataVaults Methodology was described, as deriving from the overall project's scope and envisioned functionalities, and the demonstrator input from the scenarios of D1.1. The resulting methodology consists of 8 Phases, and each Phase entails a number of Operations. The identified Phases are: Phase I: Data Retrieval, Phase II: Data Transformation & Enrichment, Phase III: Asset Storage at Individual's Side (DataVaults Personal App), Phase IV: Asset Sharing to DataVaults Cloud Platform, Phase V: DataVaults Cloud Platform Asset Storage, Phase VI: Asset Exploration & Extraction, Phase VI: Asset Exploration & Extraction, Phase VII: Data Analytics, Phase VIII: Added Value Services.

Afterwards, based on the Methodology Phases and their Operations, eleven high-level operation scenarios, as driven by the DataVaults actors, were outlined in detail. The operation scenarios were designed to abstract the demonstrators' procedures and find common ground of intervention in current procedures and bring new possibilities. The scenarios consist of a high-level description, the sequence of steps in main and alternative flows, workflow diagrams and have dedicated sections for ethical, privacy, security and GDPR-related aspects.

Finally, the DataVaults MVP was consolidated. For this purpose, a set of 79 features were derived from the operation scenarios, that were subsequently evaluated on business and technical terms by the DataVaults partners. The voting results were properly handled to extract insights and provide an initial feature prioritisation, that will drive the implementation and research planning and indicate which functionalities will be included in the DataVaults integrated platform releases.

The outcomes of this deliverable will be used as input for the definition of the DataVaults architecture and requirements in WP5 and will guide the development of the DataVaults components in WP3 and WP4.

Furthermore, the current version of the MVP, alongside with new features and new workflows, will be amended based on the input to be received by the aforementioned WPs. This will be documented in deliverable D1.4 to reflect the results of the second feature value assessment that will also involve external actors.

7 REFERENCES

[1]	Clegg, D., & Barker, R. (1994). Case method fast-track: a RAD approach. Addison-Wesley Longman Publishing Co., Inc..
[2]	Direct Anonymous Attestation - https://en.wikipedia.org/wiki/Direct_Anonymous_Attestation
[3]	Toreador Framework - http://www.toreador-project.eu
[4]	BigDL Framework - https://bigdl-project.github.io/0.11.0/
[5]	Apache Hadoop - https://hadoop.apache.org
[6]	Apache Spark - https://spark.apache.org

ANNEX I: COMPLETE RESULTS OF FEATURE VOTING

The votes of the technical feature assessment are presented in the following Table. For presentation purposes, we present here the mapping of votes to numerical values, namely 'high value = 1', 'not high value = 0', 'complex = 1', 'not complex = 0'

		VALUE												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE VALUE
DVPERS_F_01	Ability of Individual to register/login using third-party identity providers	1	0	1	1	1	1	0	1	1	1	0	1	0,75
DVPERS_F_02	Population of Individual's profile	1	1	1	1	1	1	1	1	0	1	1	1	0,92
DVPERS_F_03	AutoPopulation of Individual's profile	1	0	0	0	0	0	0	0	0	0	0	0	0,08
DVPERS_F_04	Request for advanced profile information – Individual's profile enrichment	0	0	1	1	1	0	1	1	0	1	0	1	0,58
DVPERS_F_05	Display profile completion status	0	1	0	0	1	0	0	1	0	1	0	1	0,42
DVPERS_F_06	Authorise DataVaults Personal App to collect data from local data	1	0	1	1	1	1	1	1	1	1	1	1	0,92
DVPERS_F_07	Connect Personal DataVaults App to external data sources	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPERS_F_08	Data collection configuration options	1	1	1	1	0	1	1	1	1	0	1	1	0,83
DVPERS_F_09	Test data source connection and data availability	0	0	0	1	1	0	1	1	1	1	0	1	0,58
DVPERS_F_10	Apply data quality check	0	1	1	1	1	1	1	1	1	1	1	1	0,92
DVPERS_F_11	Implement data transformations on fetched data	1	0	1	1	1	1	1	1	1	1	1	1	0,92
DVPERS_F_12	Mapping data to DataVaults schema	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPERS_F_13	Semantic annotation/enrichment of data assets	1	1	1	0	1	1	1	1	1	1	0	1	0,83
DVPERS_F_14	Data assets linking	0	1	0	0	1	1	1	0	1	1	0	1	0,58
DVPERS_F_15	Local visualisation of datasets	0	1	0	0	1	0	0	1	1	1	0	0	0,42
DVPERS_F_16	Running edge analytics	0	1	1	0	0	1	1	1	0	0	0	0	0,42
DVPERS_F_17	Generating assets from the outputs of edge analytics	0	1	0	0	0	1	1	1	0	0	0	1	0,42
DVPERS_F_18	Encryption of personal data at the Individual's side	1	1	1	1	1	1	1	1	1	1	1	0	0,92
DVPERS_F_19	Indexing of data assets at local storage	0	1	0	0	1	1	1	1	1	1	1	0	0,67
DVPERS_F_20	Storage of data assets at local storage	1	1	0	0	1	1	1	1	1	1	1	0	0,75

		VALUE												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE VALUE
DVPERS_F_21	De-authorisation of connection with given data source	1	0	1	0	1	1	1	1	1	1	1	0	0,75
DVPERS_F_22	Removal of personal data from local storage	1	1	1	1	1	1	1	1	1	1	1	0	0,92
DVPERS_F_23	Configuration of sharing anonymisation level	1	1	1	0	1	1	1	1	1	1	1	0	0,83
DVPERS_F_24	Configuration of access policies	1	1	1	0	1	1	1	1	1	1	1	0	0,83
DVPERS_F_25	Configuration of license(s)	0	0	0	0	1	1	0	1	1	1	1	0	0,50
DVPERS_F_26	*Configuration of price tag(s)*	0	1	1	1	1	1	0	0	1	1	1	1	0,75
DVPERS_F_27	Suggestion of data asset price setting	0	0	1	0	1	0	1	0	0	1	0	1	0,42
DVPERS_F_28	Assessment of sharing privacy exposure	1	1	1	1	1	1	1	1	0	1	1	1	0,92
DVPERS_F_29	Selection of DAA method at sharing by an Individual	0	0	1	0	0	1	1	1	0	0	1	0	0,42
DVPERS_F_30	Save sharing configuration	0	1	0	0	1	0	1	1	1	1	0	0	0,50
DVPERS_F_31	Use of existing sharing configuration template for a new data asset sharing job	0	1	1	0	1	0	0	1	1	1	0	0	0,50
DVPERS_F_32	Edits on saved sharing configuration	0	1	1	0	0	0	1	1	0	0	0	0	0,33
DVPERS_F_33	Generation and management of a sharing schedule	0	0	0	0	1	0	0	1	0	1	0	0	0,25
DVPERS_F_34	Management of private DataVaults contract for data asset transfer	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPERS_F_35	Individuals transaction log	1	1	1	0	1	1	1	1	1	1	1	1	0,92
DVPERS_F_36	Information on the Digital Twin Identities owned by the Individual	1	0	1	0	0	1	1	1	0	0	1	1	0,58
DVPERS_F_37	Notifications generated by DataVaults Personal App	1	1	1	0	1	0	0	1	1	1	0	1	0,67

		VALUE												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE VALUE
DVPLAT_F_01	Transferring value of shared data assets to the Individual	1	1	1	0	1	1	1	1	0	1	1	1	0,83
DVPLAT_F_02	Generation of Digital Twin	1	0	0	0	1	1	1	1	1	1	1	1	0,75
DVPLAT_F_03	Generation of Personas	1	0	1	0	1	1	1	1	1	1	1	1	0,83
DVPLAT_F_04	Enforcement of data access policies	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_05	Encryption of data assets at the DataVaults Cloud Platform	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_06	Enable searching over shared encrypted data assets	1	1	1	1	1	0	1	1	1	1	1	1	0,92
DVPLAT_F_07	Storage of data assets at DataVaults Cloud Platform	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_08	Indexing of data assets at DataVaults Cloud Platform	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_09	Removal of data assets from DataVaults Cloud Platform	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_10	Data Seekers profile editing	0	1	1	0	0	0	0	0	0	0	1	1	0,33
DVPLAT_F_11	Data Seekers profile authenticity validation	0	1	1	1	1	1	1	1	1	1	1	1	0,92
DVPLAT_F_12	Apply DAA method to Data Seeker connection	0	0	0	0	1	0	0	0	1	1	1	1	0,42
DVPLAT_F_13	Construction of simple or complex data asset queries	0	1	1	1	1	1	1	1	1	1	1	1	0,92
DVPLAT_F_14	Navigation through query results in the unencrypted data lake	1	1	0	0	1	1	1	1	1	1	1	1	0,83
DVPLAT_F_15	Navigation through query results in the encrypted data lake	1	1	1	0	1	0	1	1	1	1	1	1	0,83
DVPLAT_F_16	Custom data asset request from a Data Seeker	0	0	1	1	1	1	1	1	1	1	0	1	0,75
DVPLAT_F_17	Transferring value for acquired data assets to the DataVaults	1	0	1	0	1	1	1	1	1	1	1	1	0,83
DVPLAT_F_18	Management of public DataVaults contract for data asset acquisition	1	0	1	0	1	1	1	1	1	1	1	1	0,83
DVPLAT_F_19	Data Seekers transaction log	1	1	1	0	0	1	0	1	1	0	1	1	0,67
DVPLAT_F_20	Data Seekers assets vault	0	1	0	1	0	1	1	1	0	0	1	1	0,58

		VALUE												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE VALUE
DVPLAT_F_21	Present disk usage quotas to Data Seekers	0	0	0	0	1	0	0	0	0	1	0	1	0,25
DVPLAT_F_22	Upload own data to the Data Seekers assets vault	0	0	0	0	0	0	1	1	0	0	0	1	0,25
DVPLAT_F_23	Exporting acquired data assets from the DataVaults Cloud Platform	1	0	1	0	1	1	1	1	0	1	0	1	0,67
DVPLAT_F_24	Remote access to acquired data assets via the DataVaults Cloud Platform	0	0	0	0	0	0	0	0	1	0	1	1	0,25
DVPLAT_F_25	Present analytics time quotas to Data Seekers	0	0	1	0	1	0	0	0	0	1	0	1	0,33
DVPLAT_F_26	Configuration of data analytics task	0	0	1	0	1	1	1	1	1	1	0	0	0,58
DVPLAT_F_27	Creation of a chain of data analytics tasks	0	1	0	0	0	0	1	1	1	0	0	0	0,33
DVPLAT_F_28	Compatibility check of data assets with selected analytics task(s)	0	0	1	0	0	0	1	1	1	0	0	1	0,42
DVPLAT_F_29	Storage of data analytics task configuration	0	0	0	0	1	0	1	1	0	1	0	1	0,42
DVPLAT_F_30	Suggestion of pre-defined data analytics chains	0	1	1	0	0	1	0	0	0	0	0	1	0,33
DVPLAT_F_31	Visualisation of data analysis results on a dashboard	0	1	1	0	0	1	1	1	1	0	0	1	0,58
DVPLAT_F_32	Customisation of visualisation dashboard	0	1	0	0	0	0	0	1	0	0	0	1	0,25
DVPLAT_F_33	Storage of visualisation configuration	0	1	0	0	0	0	0	0	0	0	0	1	0,17
DVPLAT_F_34	Storage of data analysis results	0	1	1	1	1	1	1	1	1	1	0	1	0,83
DVPLAT_F_35	Storage of visualizations as projects	0	0	0	0	0	0	0	0	0	0	0	1	0,08
DVPLAT_F_36	Availability of data analysis results for download in various formats	0	1	1	1	1	1	1	1	1	1	0	1	0,83
DVPLAT_F_37	Availability of data analysis results through an API	0	1	0	1	0	1	1	1	1	0	0	1	0,58
DVPLAT_F_38	Availability of visualization results for download in various formats	0	1	0	0	0	1	0	1	0	0	0	1	0,33
DVPLAT_F_39	Complete erasure of data from the Analytics Playground	0	1	1	1	1	1	1	1	1	1	0	0	0,75
DVPLAT_F_40	*Notifications generated by DataVaults Cloud Platform*	1	1	1	0	1	0	0	1	1	1	1	0	0,67
DVPLAT_F_41	Right to be forgotten	1	1	1	1	1	1	1	1	1	1	1	0	0,92
DVPLAT_F_42	Right to data portability	1	1	1	1	1	1	1	1	1	1	1	0	0,92

		COMPLEXITY												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE COMPLEXITY
DVPERS_F_01	Ability of Individual to register/login using third-party identity providers	0	1	0	0	1	1	1	1	0	1	0	0	0,50
DVPERS_F_02	Population of Individual's profile	0	1	1	0	0	0	0	0	0	0	0	0	0,17
DVPERS_F_03	AutoPopulation of Individual's profile	1	1	0	1	0	1	0	0	0	0	1	1	0,50
DVPERS_F_04	Request for advanced profile information – Individual's profile enrichment	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_05	Display profile completion status	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_06	Authorise DataVaults Personal App to collect data from local	1	1	1	1	0	0	0	0	1	0	0	0	0,42
DVPERS_F_07	Connect Personal DataVaults App to external data sources	1	1	1	1	1	0	0	0	0	1	1	0	0,58
DVPERS_F_08	Data collection configuration options	0	0	0	0	0	1	0	0	0	0	0	0	0,08
DVPERS_F_09	Test data source connection and data availability	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_10	Apply data quality check	1	0	1	0	1	0	0	0	0	1	0	0	0,33
DVPERS_F_11	Implement data transformations on fetched data	1	0	1	0	1	0	1	1	0	1	0	1	0,58
DVPERS_F_12	Mapping data to DataVaults schema	1	0	0	1	1	1	0	1	0	1	0	1	0,58
DVPERS_F_13	Semantic annotation/enrichment of data assets	0	1	1	1	0	0	1	1	0	0	0	0	0,42
DVPERS_F_14	Data assets linking	1	1	1	1	0	1	0	0	0	0	0	1	0,50
DVPERS_F_15	Local visualisation of datasets	0	0	0	0	0	0	1	1	0	0	0	0	0,17
DVPERS_F_16	Running edge analytics	0	1	1	1	0	1	1	1	1	0	0	1	0,67
DVPERS_F_17	Generating assets from the outputs of edge analytics	0	1	1	1	1	0	1	1	1	1	0	0	0,67
DVPERS_F_18	Encryption of personal data at the Individual's side	1	1	1	0	0	0	0	0	0	0	0	0	0,25
DVPERS_F_19	Indexing of data assets at local storage	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_20	Storage of data assets at local storage	0	0	0	0	0	0	0	0	0	0	0	0	0,00

		COMPLEXITY												
Feature ID	**Feature Title**	**FOKUS**	**ATOS**	**DTU**	**UNISYSTEMS**	**IFAT**	**SUITE5**	**ASSENTiAN**	**ETA**	**UBITECH**	**IFAG**	**TECNALIA**	**MAGGIOLI**	**AVERAGE COMPLEXITY**
DVPERS_F_21	De-authorisation of connection with given data source	0	1	1	0	0	0	0	1	0	0	0	0	0,25
DVPERS_F_22	Removal of personal data from local storage	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_23	Configuration of sharing anonymisation level	1	1	1	1	0	1	1	1	0	0	0	1	0,67
DVPERS_F_24	Configuration of access policies	1	0	0	0	0	1	0	1	0	0	1	1	0,42
DVPERS_F_25	Configuration of license(s)	0	0	0	1	0	0	0	1	0	0	1	0	0,25
DVPERS_F_26	*Configuration of price tag(s)*	0	0	0	0	0	0	0	0	0	0	1	0	0,08
DVPERS_F_27	Suggestion of data asset price setting	0	1	0	1	1	1	1	1	1	1	0	1	0,75
DVPERS_F_28	Assessment of sharing privacy exposure	1	1	1	1	1	1	1	1	1	1	0	1	0,92
DVPERS_F_29	Selection of DAA method at sharing by an Individual	0	0	1	0	0	1	1	1	0	0	0	1	0,42
DVPERS_F_30	Save sharing configuration	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_31	Use of existing sharing configuration template for a new data asset sharing job	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPERS_F_32	Edits on saved sharing configuration	0	0	1	0	0	0	0	0	0	0	0	0	0,08
DVPERS_F_33	Generation and management of a sharing schedule	1	0	0	0	0	1	0	0	0	0	0	1	0,25
DVPERS_F_34	Management of private DataVaults contract for data asset transfer	1	1	1	1	1	1	1	1	0	1	1	1	0,92
DVPERS_F_35	Individuals transaction log	0	0	0	0	0	0	0	1	0	0	0	0	0,08
DVPERS_F_36	Information on the Digital Twin Identities owned by the Individual	0	1	0	1	0	0	0	1	0	0	0	0	0,25
DVPERS_F_37	Notifications generated by DataVaults Personal App	0	0	0	0	1	0	0	0	0	1	0	0	0,17

		COMPLEXITY												
Feature ID	**Feature Title**	**FOKUS**	**ATOS**	**DTU**	**UNISYSTEMS**	**IFAT**	**SUITE5**	**ASSENTiAN**	**ETA**	**UBITECH**	**IFAG**	**TECNALIA**	**MAGGIOLI**	**AVERAGE COMPLEXITY**
DVPLAT_F_01	Transferring value of shared data assets to the Individual	1	1	1	0	1	1	1	1	0	1	0	1	0,75
DVPLAT_F_02	Generation of Digital Twin	0	0	0	1	1	1	1	1	1	1	1	1	0,75
DVPLAT_F_03	Generation of Personas	0	0	0	0	1	1	1	1	1	1	1	1	0,67
DVPLAT_F_04	Enforcement of data access policies	1	1	1	0	1	1	0	1	1	1	1	1	0,83
DVPLAT_F_05	Encryption of data assets at the DataVaults Cloud Platform	0	0	1	0	1	1	1	1	1	1	0	1	0,67
DVPLAT_F_06	Enable searching over shared encrypted data assets	1	1	1	1	1	1	1	1	1	1	1	1	1,00
DVPLAT_F_07	Storage of data assets at DataVaults Cloud Platform	0	0	0	0	1	0	0	0	0	1	0	0	0,17
DVPLAT_F_08	Indexing of data assets at DataVaults Cloud Platform	0	0	1	0	1	0	1	1	0	1	0	0	0,42
DVPLAT_F_09	Removal of data assets from DataVaults Cloud Platform	0	0	1	0	0	0	0	0	0	0	0	0	0,08
DVPLAT_F_10	Data Seekers profile editing	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPLAT_F_11	Data Seekers profile authenticity validation	0	1	0	0	0	0	1	1	1	0	1	0	0,42
DVPLAT_F_12	Apply DAA method to Data Seeker connection	0	0	1	0	1	1	1	1	1	1	0	1	0,67
DVPLAT_F_13	Construction of simple or complex data asset queries	0	1	1	1	0	0	0	0	0	0	0	0	0,25
DVPLAT_F_14	Navigation through query results in the unencrypted data lake	0	1	0	0	0	0	0	0	0	0	0	0	0,08
DVPLAT_F_15	Navigation through query results in the encrypted data lake	1	1	1	0	0	1	1	1	1	0	0	1	0,67
DVPLAT_F_16	Custom data asset request from a Data Seeker	1	1	1	0	0	1	0	0	0	0	1	1	0,50
DVPLAT_F_17	Transferring value for acquired data assets to the DataVaults	0	1	0	0	0	1	1	1	0	0	0	1	0,42
DVPLAT_F_18	Management of public DataVaults contract for data asset acquisition	1	0	1	1	1	1	1	1	0	1	0	1	0,75
DVPLAT_F_19	Data Seekers transaction log	0	1	0	0	0	0	0	0	0	0	0	0	0,08
DVPLAT_F_20	Data Seekers assets vault	0	0	0	0	0	0	1	1	0	0	0	0	0,17

		COMPLEXITY												
Feature ID	Feature Title	FOKUS	ATOS	DTU	UNISYSTEMS	IFAT	SUITE5	ASSENTiAN	ETA	UBITECH	IFAG	TECNALIA	MAGGIOLI	AVERAGE COMPLEXITY
DVPLAT_F_21	Present disk usage quotas to Data Seekers	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPLAT_F_22	Upload own data to the Data Seekers assets vault	1	0	1	0	1	0	0	0	0	1	0	0	0,33
DVPLAT_F_23	Exporting acquired data assets from the DataVaults Cloud Platform	1	0	0	1	0	0	0	0	0	0	0	0	0,17
DVPLAT_F_24	Remote access to acquired data assets via the DataVaults Cloud Platform	0	0	1	1	1	0	1	0	0	1	0	0	0,42
DVPLAT_F_25	Present analytics time quotas to Data Seekers	0	0	0	0	0	1	1	1	1	0	0	1	0,42
DVPLAT_F_26	Configuration of data analytics task	0	0	1	0	1	1	1	1	0	1	0	1	0,58
DVPLAT_F_27	Creation of a chain of data analytics tasks	0	1	1	0	0	1	1	1	1	0	0	1	0,58
DVPLAT_F_28	Compatibility check of data assets with selected analytics task(s)	0	0	1	0	1	1	0	1	1	1	0	1	0,58
DVPLAT_F_29	Storage of data analytics task configuration	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPLAT_F_30	Suggestion of pre-defined data analytics chains	0	0	0	0	1	1	1	1	0	1	0	1	0,50
DVPLAT_F_31	Visualisation of data analysis results on a dashboard	0	0	0	0	0	0	1	1	0	0	0	0	0,17
DVPLAT_F_32	Customisation of visualisation dashboard	0	1	0	0	0	1	0	1	0	0	0	1	0,33
DVPLAT_F_33	Storage of visualisation configuration	0	0	0	0	0	0	0	1	0	0	0	0	0,08
DVPLAT_F_34	Storage of data analysis results	0	0	0	0	0	0	1	1	0	0	0	0	0,17
DVPLAT_F_35	Storage of visualizations as projects	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPLAT_F_36	Availability of data analysis results for download in various formats	0	0	0	0	0	0	0	0	0	0	0	0	0,00
DVPLAT_F_37	Availability of data analysis results through an API	0	0	1	0	1	0	0	0	0	1	0	0	0,25
DVPLAT_F_38	Availability of visualization results for download in various formats	0	0	0	0	0	0	1	1	0	0	0	0	0,17
DVPLAT_F_39	Complete erasure of data from the Analytics Playground	0	1	1	0	0	0	1	1	0	0	0	0	0,33
DVPLAT_F_40	*Notifications generated by DataVaults Cloud Platform*	0	0	0	0	1	0	0	0	0	1	0	0	0,17
DVPLAT_F_41	Right to be forgotten	0	1	1	0	1	0	1	1	0	1	1	0	0,58
DVPLAT_F_42	Right to data portability	0	1	1	1	1	0	1	1	0	1	1	0	0,67

The votes of the demonstrator feature assessment are presented in the following Table. For presentation purposes, we present here the mapping of votes to numerical values, namely 'Won't-have'= 1, ''Could-have'= 2', 'Should-have'= 3 and 'Must-have'= 4

Feature ID	Feature Title	OLYMPIACOS	PIRAEUS	ANDAMAN7	MIWENERGIA	PRATO	AVERAGE VALUE	MoSCoW LABEL
DVPERS_F_01	Ability of Individual to register/login using third-party identity providers	3	3	2	2	2	**2,4**	COULD HAVE
DVPERS_F_02	Population of Individual's profile	4	4	4	4	4	**4**	MUST HAVE
DVPERS_F_03	AutoPopulation of Individual's profile	3	3	3	3	4	**3,2**	SHOULD HAVE
DVPERS_F_04	Request for advanced profile information – Individual's profile enrichment	4	2	1	2	4	**2,6**	COULD HAVE
DVPERS_F_05	Display profile completion status	2	2	3	2	4	**2,6**	COULD HAVE
DVPERS_F_06	Authorise DataVaults Personal App to collect data from local data sources	4	3	3	4	4	**3,6**	SHOULD HAVE
DVPERS_F_07	Connect Personal DataVaults App to external data sources	4	3	4	4	4	**3,8**	SHOULD HAVE
DVPERS_F_08	Data collection configuration options	3	4	4	3	4	**3,6**	SHOULD HAVE
DVPERS_F_09	Test data source connection and data availability	2	3	3	3	3	**2,8**	COULD HAVE
DVPERS_F_10	Apply data quality check	3	3	3	3	3	**3**	SHOULD HAVE
DVPERS_F_11	Implement data transformations on fetched data	3	3	3	2	3	**2,8**	COULD HAVE
DVPERS_F_12	Mapping data to DataVaults schema	4	4	4	1	4	**3,4**	SHOULD HAVE
DVPERS_F_13	Semantic annotation/enrichment of data assets	2	2	3	2	4	**2,6**	COULD HAVE
DVPERS_F_14	Data assets linking	2	3	3	2	2	**2,4**	COULD HAVE
DVPERS_F_15	Local visualisation of datasets	3	2	3	3	4	**3**	SHOULD HAVE
DVPERS_F_16	Running edge analytics	3	2	2	2	3	**2,4**	COULD HAVE
DVPERS_F_17	Generating assets from the outputs of edge analytics	3	2	2	3	2	**2,4**	COULD HAVE
DVPERS_F_18	Encryption of personal data at the Individual's side	2	3	4	3	4	**3,2**	SHOULD HAVE
DVPERS_F_19	Indexing of data assets at local storage	2	2	3	4	4	**3**	SHOULD HAVE
DVPERS_F_20	Storage of data assets at local storage	3	2	4	4	4	**3,4**	SHOULD HAVE

Feature ID	Feature Title	OLYMPIACOS	PIRAEUS	ANDAMAN7	MIWENERGIA	PRATO	AVERAGE VALUE	MoSCoW LABEL
DVPERS_F_21	De-authorisation of connection with given data source	4	3	4	4	4	**3,8**	SHOULD HAVE
DVPERS_F_22	Removal of personal data from local storage	4	4	4	4	4	**4**	MUST HAVE
DVPERS_F_23	Configuration of sharing anonymisation level	3	4	4	4	4	**3,8**	SHOULD HAVE
DVPERS_F_24	Configuration of access policies	4	3	4	3	4	**3,6**	SHOULD HAVE
DVPERS_F_25	Configuration of license(s)	3	2	3	3	4	**3**	SHOULD HAVE
DVPERS_F_26	Configuration of price tag(s)	4	3	3	3	4	**3,4**	SHOULD HAVE
DVPERS_F_27	Suggestion of data asset price setting	2	3	2	2	3	**2,4**	COULD HAVE
DVPERS_F_28	Assessment of sharing privacy exposure	2	4	2	3	3	**2,8**	COULD HAVE
DVPERS_F_29	Selection of DAA method at sharing by an Individual	2	2	3	3	4	**2,8**	COULD HAVE
DVPERS_F_30	Save sharing configuration	2	2	3	2	4	**2,6**	COULD HAVE
DVPERS_F_31	Use of existing sharing configuration template for a new data asset sharing job	2	2	2	2	3	**2,2**	COULD HAVE
DVPERS_F_32	Edits on saved sharing configuration	2	2	1	2	2	**1,8**	COULD HAVE
DVPERS_F_33	Generation and management of a sharing schedule	2	2	3	2	4	**2,6**	COULD HAVE
DVPERS_F_34.	Management of private DataVaults contract for data asset transfer	2	4	4	4	4	**3,6**	SHOULD HAVE
DVPERS_F_35	Individuals transaction log	4	2	3	3	4	**3,2**	SHOULD HAVE
DVPERS_F_36	Information on the Digital Twin Identities owned by the Individual	3	2	3	3	4	**3**	SHOULD HAVE
DVPERS_F_37	Notifications generated by DataVaults Personal App	3	3	3	2	4	**3**	SHOULD HAVE

Feature ID	Feature Title	OLYMPIACOS	PIRAEUS	ANDAMAN7	MIWENERGIA	PRATO	AVERAGE VALUE	MoSCoW LABEL
DVPLAT_F_01	Transferring value of shared data assets to the Individual	4	3	4	3	4	**3,6**	SHOULD HAVE
DVPLAT_F_02	Generation of Digital Twin	2	2	3	3	4	**2,8**	COULD HAVE
DVPLAT_F_03	Generation of Personas	2	2	2	2	4	**2,4**	COULD HAVE
DVPLAT_F_04	Enforcement of data access policies	4	4	4	4	4	**4**	MUST HAVE
DVPLAT_F_05	Encryption of data assets at the DataVaults Cloud Platform	4	4	4	3	4	**3,8**	SHOULD HAVE
DVPLAT_F_06	Enable searching over shared encrypted data assets	2	3	3	2	4	**2,8**	COULD HAVE
DVPLAT_F_07	Storage of data assets at DataVaults Cloud Platform	4	3	4	2	4	**3,4**	SHOULD HAVE
DVPLAT_F_08	Indexing of data assets at DataVaults Cloud Platform	2	3	3	2	4	**2,8**	COULD HAVE
DVPLAT_F_09	Removal of data assets from DataVaults Cloud Platform	4	3	4	3	4	**3,6**	SHOULD HAVE
DVPLAT_F_10	Data Seekers profile editing	3	3	4	4	4	**3,6**	SHOULD HAVE
DVPLAT_F_11	Data Seekers profile authenticity validation	1	4	3	4	4	**3,2**	SHOULD HAVE
DVPLAT_F_12	Apply DAA method to Data Seeker connection	2	3	3	3	4	**3**	SHOULD HAVE
DVPLAT_F_13	Construction of simple or complex data asset queries	3	4	4	4	4	**3,8**	SHOULD HAVE
DVPLAT_F_14	Navigation through query results in the unencrypted data lake	4	3	4	4	4	**3,8**	SHOULD HAVE
DVPLAT_F_15	Navigation through query results in the encrypted data lake	2	3	3	3	4	**3**	SHOULD HAVE
DVPLAT_F_16.	Custom data asset request from a Data Seeker	3	2	3	2	4	**2,8**	COULD HAVE
DVPLAT_F_17	Transferring value for acquired data assets to the DataVaults Platform	4	3	3	2	4	**3,2**	SHOULD HAVE
DVPLAT_F_18	Management of public DataVaults contract for data asset acquisition	4	3	3	3	4	**3,4**	SHOULD HAVE
DVPLAT_F_19	Data Seekers transaction log	4	2	3	3	4	**3,2**	SHOULD HAVE
DVPLAT_F_20	Data Seekers assets vault	4	3	3	2	2	**2,8**	COULD HAVE
DVPLAT_F_21	Present disk usage quotas to Data Seekers	2	2	3	2	2	**2,2**	COULD HAVE
DVPLAT_F_22	Upload own data to the Data Seekers assets vault	1	3	3	2	2	**2,2**	COULD HAVE
DVPLAT_F_23	Exporting acquired data assets from the DataVaults Cloud Platform	4	4	2	3	4	**3,4**	SHOULD HAVE
DVPLAT_F_24	Remote access to acquired data assets via the DataVaults Cloud Platform	2	3	4	3	4	**3,2**	SHOULD HAVE
DVPLAT_F_25	Present analytics time quotas to Data Seekers	3	2	3	1	2	**2,2**	COULD HAVE
DVPLAT_F_26	Configuration of data analytics task	2	2	3	3	4	**2,8**	COULD HAVE
DVPLAT_F_27	Creation of a chain of data analytics tasks	2	2	3	1	3	**2,2**	COULD HAVE
DVPLAT_F_28	Compatibility check of data assets with selected analytics task(s)	2	2	4	4	4	**3,2**	SHOULD HAVE
DVPLAT_F_29	Storage of data analytics task configuration	2	2	3	2	4	**2,6**	COULD HAVE
DVPLAT_F_30	Suggestion of pre-defined data analytics chains	2	2	2	1	3	**2**	COULD HAVE
DVPLAT_F_31	Visualisation of data analysis results on a dashboard	3	2	3	3	4	**3**	SHOULD HAVE
DVPLAT_F_32	Customisation of visualisation dashboard	2	2	3	2	4	**2,6**	COULD HAVE
DVPLAT_F_33	Storage of visualisation configuration	2	2	3	1	4	**2,4**	COULD HAVE
DVPLAT_F_34	Storage of data analysis results	3	2	3	3	4	**3**	SHOULD HAVE
DVPLAT_F_35	Storage of visualizations as projects	2	2	2	1	3	**2**	COULD HAVE
DVPLAT_F_36	Availability of data analysis results for download in various formats	3	2	3	2	4	**2,8**	COULD HAVE
DVPLAT_F_37	Availability of data analysis results through an API	2	2	3	1	3	**2,2**	COULD HAVE
DVPLAT_F_38	Availability of visualization results for download in various formats	2	2	3	2	4	**2,6**	COULD HAVE
DVPLAT_F_39	Complete erasure of data from the Analytics Playground	4	2	3	4	4	**3,4**	SHOULD HAVE
DVPLAT_F_40	Notifications generated by DataVaults Cloud Platform	3	3	3	2	4	**3**	SHOULD HAVE
DVPLAT_F_41	Right to be forgotten	4	4	4	4	4	**4**	MUST HAVE
DVPLAT_F_42	Right to data portability	4	4	4	4	4	**4**	MUST HAVE

ANNEX II: DEMONSTRATION SCENARIOS – OPERATION SCENARIOS MAPPING

Scenario #	OLIMPIAKOS Members	OLIMPIAKOS Athletes	Piraeus – Mobility	Piraeus - Entrepreneurs	Piraeus – Culture/Tourism	MIWENERGIA – PV Installation	MIWENERGIA – Energy	MIWENERGIA Energy Trends	Andaman7 - Cloud	Andaman7 - Smartphone	Prato - Mobility	Prato - Culture/Tourism	Prato – Civil Certificates
1. Personal Data Collection. An Individual uses the Personal App to:	X	X	X	X	X	X	X	X	X	X	X	X	X
Manually update profile data		X	X	X	X	X	X	X	X	X	X	X	X
Collect personal data from a specific data source	X			X		X	X	X	X	X	X	X	X
Requirements that are different from the ones described.													
A. Olimpiakos Members	Regarding the members' scenario, an interconnection between DataVaults platform and the club CRM is required.												
I Andaman7 - Cloud	Concern data already stored in Andaman7 app												
J. Andaman7 - Smartphone	Concern data coming from various source like hospitals												
M Prato – Civil Certificates	At workflow reference 3.b.i. Create connection with a webservice to collect data												
2: Personal Data Assets Exploration & Analysis An Individual browses and experiments with their data.	X	X				X	X	X		X	X	X	X
An Individual uses the App to browse through his/her own data, create simple analyses, visualise them and get a holistic picture of which data assets the Individual has shared over DataVaults.	X	X				X	X	X		X	X	X	X
Requirements that are different from the ones described.													
M Prato – Civil Certificates	At workflow reference 4.c. [Automatic] deletion/update of obsolete data (e.g. expired certificates)												
3: Personal Data Assets Sharing Gains and Risk Information An Individual is informed about earnings and risk exposure.	X	X	X	X	X	X	X	X	X		X	X	X
An Individual uses DataVaults Personal App to learn about the transactions he already completed over	X	X	X	X	X	X	X	X	X		X	X	X

DataVaults and current privacy risk exposure.													
Requirements that are different from the ones described.													
4. Personal Data Assets Sharing Configuration	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
An individual creates a configuration for the sharing of their data assets on the DataVaults Cloud Platform (data selection, anonymisation mechanisms, access policies design, asset value description, etc.)	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
Requirements that are different from the ones described.													
5. Personal Data Sharing / Cloud Upload	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
An Individual selects to share personal data with DataVaults, has already set sharing configuration parameters (sharing configuration already set through Scenario 4: Personal Data Assets Sharing Configuration), receives a success notification and updated metrics on privacy risk exposure.	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
Requirements that are different from the ones described.													
6. Personal Data Assets Sharing Revocation	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
The Individual selects one data asset already shared over the DataVaults Cloud Platform and chooses to revoke sharing. The individual receives a success notification and updated metrics on privacy risk exposure.	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**	**X**
Requirements that are different from the ones described.													
7. Explore Data Assets	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
A Data Seeker connects to DataVaults Cloud Platform through different interfaces and explores anonymised data using search mechanisms such as the DataVaults Data Lake, in	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	

order to locate data which s/he is interested in.													
A Data Seeker connects to DataVaults Cloud Platform through different interfaces and explores unmasked (eponymous) data from Individuals.	**X**	**X**				**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Requirements that are different from the ones described.													
MIWENERGIA F,G and H	We consider that the data providers are the ones who choose to share unmasked data, or let the seekers explore it before the trade. As in many usage scenario "a data seeker uses DataVaults, request data, connect to" or "an individual selects to, create a configuration etc" we think that the DataVaults users are who have to select some of the options in their account´s settings (privacy or wherever) for usage scenarios like number 6, 7 and 8. The availability of data for data seekers should be defined by the data providers.												
ANDAMAN7	Note (also true for following features) : **I.** Dataseeker is hospitals, clinical trial, research but also the app (for backup). J. Dataseeker is the Andaman7 app that will collect personal data that are not yet in the app but can also aggregated data from other patients with the same profile As health partners are difficult to convince, those aspects will be harder to get in the context of scenario I than in scenario J (where Andaman7 is the data seeker)												
8. Acquire Data Assets from the DataVaults Cloud	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
A Data Seeker requests data assets s/he has already located (e.g. through exploration) and offers the listed compensation for those by issuing a smart contract	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Requirements that are different from the ones described.													
ANDAMAN7	As health partners are difficult to convince, those aspects will be harder to get in the context of scenario I than in scenario J (where Andaman7 is the data seeker)												
9. Acquire Data Assets from a DataVaults Individual User	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
A Data Seeker is already connected and wants to acquire data directly from Individuals through the DataVaults platform. The Individual is informed and chooses to accept the compensation, or not.	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Requirements that are different from the ones described.													

ANDAMAN7	As health partners are difficult to convince, those aspects will be harder to get in the context of scenario I than in scenario J (where Andaman7 is the data seeker)												
PRATO K,L,M	Alt. 7. In case obsolete data have been deleted (scenario 2.4.c) activate scenario 1 to collect new data not yet available at the Individual's side.												
10. Analyse and Visualise Data	**X**	**X**				**X**	**X**	**X**	**X**	**X**	**X**	**X**	
A Data Seeker uses DataVaults Cloud Platform to analyse personal data s/he has already acquired through DataVaults Cloud Platform, possibly combine them with personal data s/he has already in his/her possession and visualise the results in order to extract insights.	**X**	**X**				**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Requirements that are different from the ones described.													
PIRAEUS	In all scenarios analysis and visualization is expected to happen through the use of some other software												
ANDAMAN7	As health partners are difficult to convince, those aspects will be harder to get in the context of scenario I than in scenario J (where Andaman7 is the data seeker)												
11. Create Ready-Made Analysis of Assets available in the DataVaults Cloud Platform.	**X**	**X**				**X**	**X**	**X**	**X**	**X**	**X**	**X**	
A DataVaults Data Scientist connects to DataVaults Cloud Platform in order to analyse data and create ready-made analyses for others (e.g. Data Seekers) to find available upon connection.	**X**	**X**				**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Requirements that are different from the ones described.													
ANDAMAN7	As health partners are difficult to convince, those aspects will be harder to get in the context of scenario I than in scenario J (where Andaman7 is the data seeker)												
OTHER REQUIREMENTS FROM SCENARIOS WHICH MAY HAVE BEEN OVERLOOKED AND ARE DEEMED IMPORTANT.													
PRATO: Collection of dynamic personal data through a pop up asking for info (e.g. "which means of transportation are you using now?")													
ANDAMAN7 The deliverable talks about data in general. I can see that there is chapter to talk about data and it will probably details what kind of data is supported. The medical field is evolving and we have more and more structured and standardize data but there are still a lot of « bad » files in the wild (old PDFs, scan image, …) that add difficulty to what we are trying to do and that should definitely be considered.													