DataVaults

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

D3.3
Security, Privacy and Trust Bundles - Version 3

Editor(s)	Alexander Köberl
Lead Beneficiary	Infineon Technologies Austria AG (IFAT)
Status	Final
Version	v1.00
Due Date	31/08/2022
Delivery Date	10/10/2022
Dissemination Level	PU

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

Project	DataVaults – 871755
Work Package	WP3 – Bundles for Secure Data Sharing and Access, Privacy and Trust Preservation and IPRs Management
Deliverable	D3.3 – Security, Privacy and Trust Bundles - Version 3
Editor(s)	IFAT – Alexander Köberl
Contributor(s)	Assentian – Ilesh Dattani, ATOS – Raquel Cortés Carreras, IFAT – Holger Bock, Suite5 – Sotiris Koussouris Tecnalia – María Jose Lopez, UBITECH – Giannis Ledakis UNISYSTEMS – John Kaldis
Reviewer(s)	DTU – Weizhi Meng Suite5 – Sotiris Koussouris

Abstract	This deliverable represents an auxiliary document, describing the progress of the code and component development under WP3.
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

Version	Date	Partner	Description
0.10	14/07/2022	IFAT	Updated document
0.20	01/08/2022	IFAT	Merged versions from Tecnalia, Assentian, ATOS, Suite5, IFAT
0.30	05/08/2022	IFAT	Corrections by IFAT
0.90	31/08/2022	IFAT	Merged update by UBITECH
0.91	09/09/2022	DTU/Suite5	Review
0.92	28/09/2022	Suite5	Screenshots for Anonymizer, Persona Generator
0.93	06/10/2022	ATOS/Suite5	Screenshots for Personal Wallet
0.94	07/10/2022	Ubitech	Additions to DLT Engine
1.00	10/10/2022	IFAT / Suite5	Final Version to be Submitted to the EC

Executive Summary

This deliverable represents an auxiliary document to summarize the progress of code development in the individual components for both the Personal App and Cloud Platform of DataVaults. It collects the direct outcome of all WP3 tasks in a deliverable of type OTHER, which was updated by three versions. The intermediate results reflect the development progress made in months 25-32 within this work package. This is the final deliverable of this work package.

The components that are presented in this deliverable are split into DataVaults Personal App and Cloud Platform. This is done to group related components and clearly assign the responsibility to the target platform.

Personal App:

- **Trusted Platform Module (TPM) Interface** to provide a hardware anchor for privacy and attestation services.
- **Sharing configurator** to provide a user interface to collect the data and apply the selected sharing operations.
- **Data Request Service Resolver** for receiving asynchronous inquiries for additional data and updated sharing preferences.
- **Personal Wallet** for managing the compensation in a privacy-preserving manner.
- **Blockchain Security 2Go Starter Kit** as the interface to secure key management and transaction signing with smart cards.
- **Data Anonymizer** for obfuscating the data in order to preserve the privacy of the user.
- **Attribute Based Encryption Engine** to ensure the confidentiality of the data.
- **Access Policies Editor** for configuring conditions for data sharing.

Cloud Platform:

- **Access Policy Engine** to control the access to specific data.
- **Risk Management Monitor and Dashboard** to monitor and evaluate the risks related to the privacy exposure.
- **DataStream and Contract Composer** to manage the lifecycle of the contracts.
- **Policy-compliant Blockchain Infrastructure and DLT Engine** to facilitate the sealing of contracts on the side of the Individuals, as well as their compensation for assets that have been bought by Data Seekers.
- **Persona Generator** to support the data anonymization process.

Table of Contents

List of Figures

Terms and Abbreviations

ABE	Attribute based encryption
API	Application Programming Interface
CIV	Configuration Integrity Verification
DAA	Direct Anonymous Attestation
DLT	Distributed Ledger Technology
NFC	Near Field Communication
PII	Personal Identifiable Information
REST	Representational State Transfer
TPM	Trusted Platform Module
UI	User Interface
UML	Unified Modelling Language
WP	Work Package

1. INTRODUCTION

This document represents an auxiliary document to deliver the code and give the implementation status of the components grouped under WP3. The theoretical foundation elaborated in WP2, in respect to technical, legal, and ethical requirements, is put into practice with this work package.

It is noted that for reasons of completeness, that in case the scope and the technology background of a component has not been changed, the texts describing those in this section and in the sections following are essentially the same as those presented in D3.2 [1]. In case there are changes, these are marked under a subsection in each description.

1.1 DOCUMENT STRUCTURE

After the introduction in Section 1, the document continues with the description of the components of the DataVaults Personal App in Section 2 and the corresponding documentation of the DataVaults Cloud Platform in Section 3. It gives an overview and highlights the progress until M32 of the project runtime.

A short description of each component is enriched with mock-up illustrations or real-life screenshots (if it has a User Interface (UI) and depending on the development progress) to give an impression about the intended usage from the user's point of view. The technical details include the technology stack providing the foundation of the component.

For each component, the corresponding user stories coming out of WP5 are collected and a classification between already implemented and future extensions is performed. The features are described in the form "As a *<Role>*, I want to *<Action>*, so that *<Reason>*". This gives a clear indication about the required features from the individual stakeholders' views to guide the development. All user stories are collected in tables and assigned to components to record a backlog for additional features in upcoming releases.

Finally, Section 4 concludes this document.

1.2 RELATION TO OTHER WPS/TASKS

D3.3 is the third in a series of deliverables as part of WP3 activities, with close relations to WP5 and WP4. The overall system architecture with the classification in individual components, as well as the user stories describing the target functionalities, are directly integrated as input from WP5. Moreover, results from WP3 will be returned to WP5 activities for testing and integration into the overall DataVaults solution.

The DataVaults platform is developed in three self-contained iterations, extending the supported functionality with each step. This agile development approach, where the end-product is developed in consecutive iterations, allows for flexible consideration of new findings and design choices, which are also carried over to the architecture definition and user stories if changes are required.

2 WP3 Components Descriptions – Personal App

The DataVaults Personal App is the integrated application that resides at the end of the Data Owners and is used by these users to collect, store, process and share their data with Data Seekers. All the data is handled based on specific sharing configurations by the DataVaults Cloud-based Engine. Those configurations are stored on a Blockchain and record the user's choice for the license terms, anonymization degree and many more fine-grained settings.

The majority of the components in the DataVaults Personal App are responsible for handling the data immediately after it is collected.

As such, this subsection provides the progress done in the following components of the DataVaults Personal App:

- Trusted Platform Module
- Sharing Configurator
- Data Request Service Resolver
- Personal Wallet
- Blockchain Security 2GO Starter Kit
- Data Anonymiser
- Attribute Based Encryption Engine
- Access Policies Editor

The source code of the different components, which are open source, is provided in the following repository

https://www.gitlab.com/DataVaults

It is noted that for reasons of completeness in case the scope and the technology background of a component have not been changed, the texts describing in the following sections are essentially the same as those presented in D3.2. In case there are changes, these are clearly marked in each description.

2.1 TRUSTED PLATFORM MODULE (TPM) INTERFACE

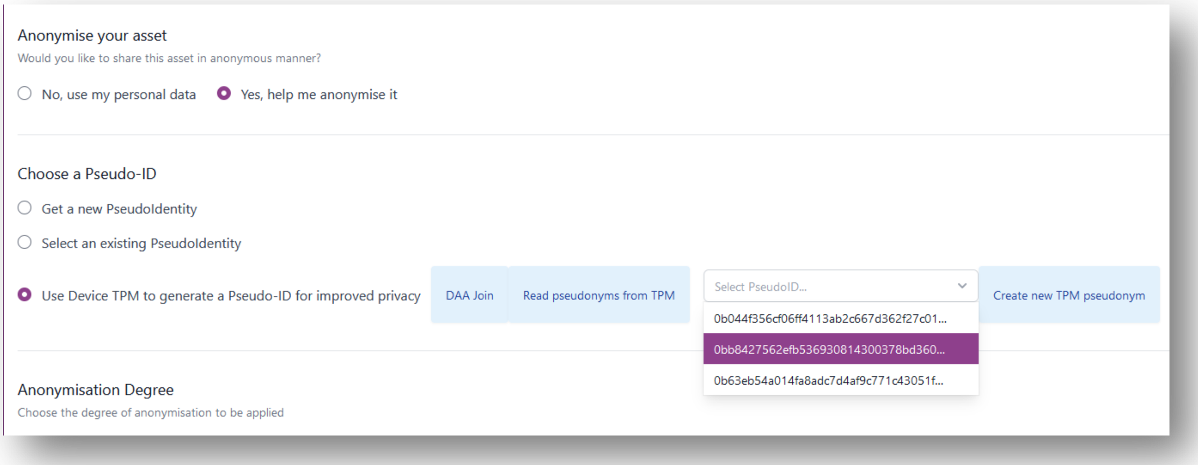
Component's Concept Update from D3.2:

As decided by the consortium, the current and future version of DataVaults will be solely offering to Data Owners an infrastructure (DataVaults Personal App) which is provisioned as a cloud-based service, accessed through their devices as a Web App. In this context, the introduction of Configuration Integrity Verification (CIV) does not offer any advantage to the current version of the Personal App, as the hardware TPM necessary for it to operate will be in each case the same (e.g. of the cloud based server, as there is no native PersonalApp client running on user's devices).

Nevertheless, the consortium has worked (as indicated in the WP3 deliverables) towards establishing the necessary methods for CIV to work and has tested it in lab environment, in order to facilitate support of this optional operation of the DataVaults in case native applications are developed in the post-project exploitation period.

Component's Description: TPM is a low-level hardware component, which offers a variety of features to enhance the security and trust characteristics of an application. In DataVaults, it enables Direct Anonymous Attestation (DAA): Uploaded data can optionally be signed with this privacy-preserving scheme. This allows the Data Seeker to receive additional certification of the data and verify that it originated from a group of certified users, while protecting the identity of the individual user.

This scheme requires additional software installed and access to the TPM on the client device. For the purely browser-based access delivered to the demonstrators, this is not possible. Nevertheless, definition and implementation of the interfaces for seamless integration is done and verified by the technical partners. The TPM interface does not have an exclusive user interface; it is instead part of other components (e.g. sharing configurator).



Anonymise your asset

Would you like to share this asset in anonymous manner?

○ No, use my personal data ● Yes, help me anonymise it

Choose a Pseudo-ID

○ Get a new PseudoIdentity

○ Select an existing PseudoIdentity

● Use Device TPM to generate a Pseudo-ID for improved privacy

0b044f356cf06ff4113ab2c667d362f27c01...

0bb8427562efb536930814300378bd360...

0b63eb54a014fa8adc7d4af9c771c43051f...

Anonymisation Degree

Choose the degree of anonymisation to be applied

Figure 1: TPM interface integrated in the Sharing Configurator

2.1.1 Technology Background

<u>Technology Background Update from D3.2:</u> None

Technology Description: For the DAA scheme, we use the already available implementation written in C++. Internally they rely on the IBM TPM Software Stack[1] (TSS) for interaction with the TPM API and OpenSSL for generic cryptographic operations.

The interface is implemented as a local socket in *Python 3* to provide the API to other components, pre-process input data and forward calls to the responsible service.

2.1.2 Component Backlog

2.1.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID#	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_059	Sharing Configurator. Sharing Setup Manager	Data Provider	Select if I want DAA to be used during data sharing	my data asset will have an extra “trustworthiness” guarantee.
US_068	Attestation	Data Provider	Authenticate my device in a privacy-preserving way	my data cannot be linked directly to my identity.
US_069	Attestation	Data Provider	I want to get informed if my device supports DAA	I am aware of whether the strictest privacy-preserving mode is available for me
US_070	Attestation	Data Provider	choose an alternative authentication method, in case my device does not support DAA	I have compatibility with my devices and avoid technology lock-in.
US_071	Attestation	Data Provider	have the strictest privacy-preserving technology enabled by default	I have the best privacy preservation without further configuration.

2.1.2.2 Features planned for upcoming Releases

All identified features of the initial platform requirements have been delivered.

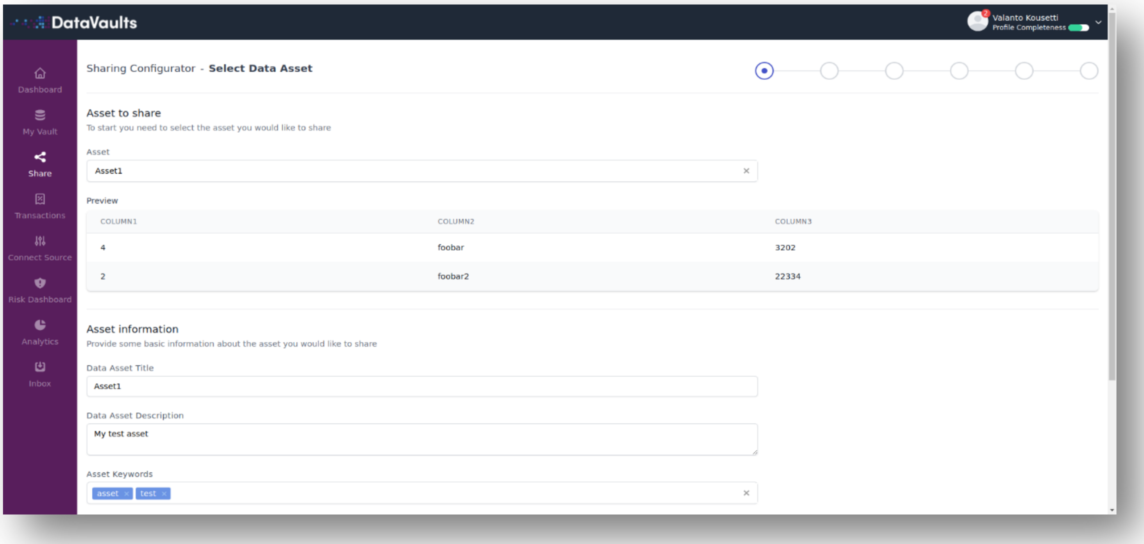
[1] https://sourceforge.net/projects/ibmtpm20tss/

2.2 SHARING CONFIGURATOR

<u>Component's Concept Update from D3.2:</u> None

Component's Description: The sharing configuration is one of the core components of the DataVaults Personal App, as it is responsible for sharing the data of a Data Owner with interested stakeholders, by putting that data on display (setting different access/security and privacy policies) over the DataVaults Cloud Platform.

Through this component, a user is able to select a data source (Figure 2) for sharing and provide extra details, then decide on whether it should be anonymized (and how) or not (see Section 2.6 and Figure 3).



Sharing Configurator - **Select Data Asset**

Asset to share

To start you need to select the asset you would like to share

Asset: Asset1

Preview

COLUMN1	COLUMN2	COLUMN3
4	foobar	3202
2	foobar2	22334

Asset information

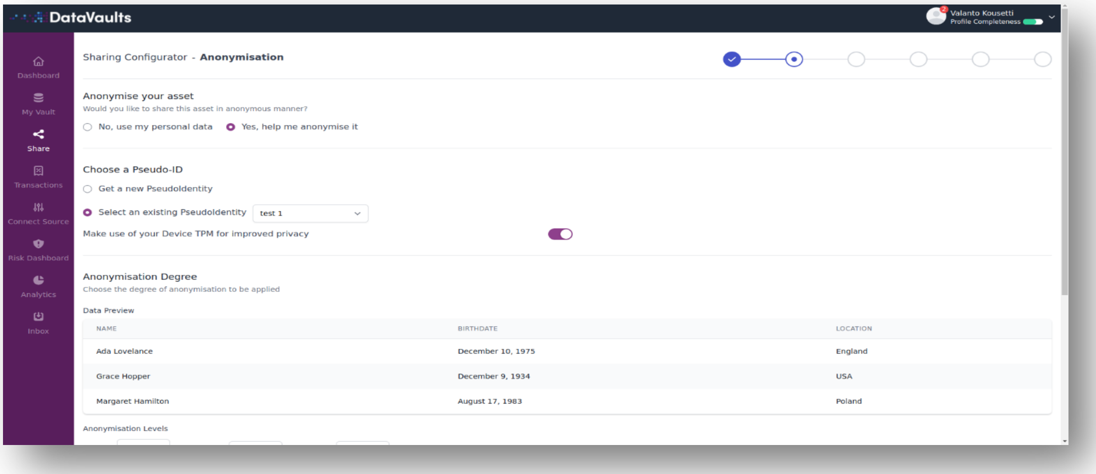
Provide some basic information about the asset you would like to share

Data Asset Title: Asset1

Data Asset Description: My test asset

Asset Keywords: asset × test ×

Figure 2: Sharing Configurator - Asset Selection for Sharing



Sharing Configurator - **Anonymisation**

Anonymise your asset

Would you like to share this asset in anonymous manner?

○ No, use my personal data ● Yes, help me anonymise it

Choose a Pseudo-ID

○ Get a new PseudoIdentity

● Select an existing PseudoIdentity: test 1

Make use of your Device TPM for improved privacy

Anonymisation Degree

Choose the degree of anonymisation to be applied

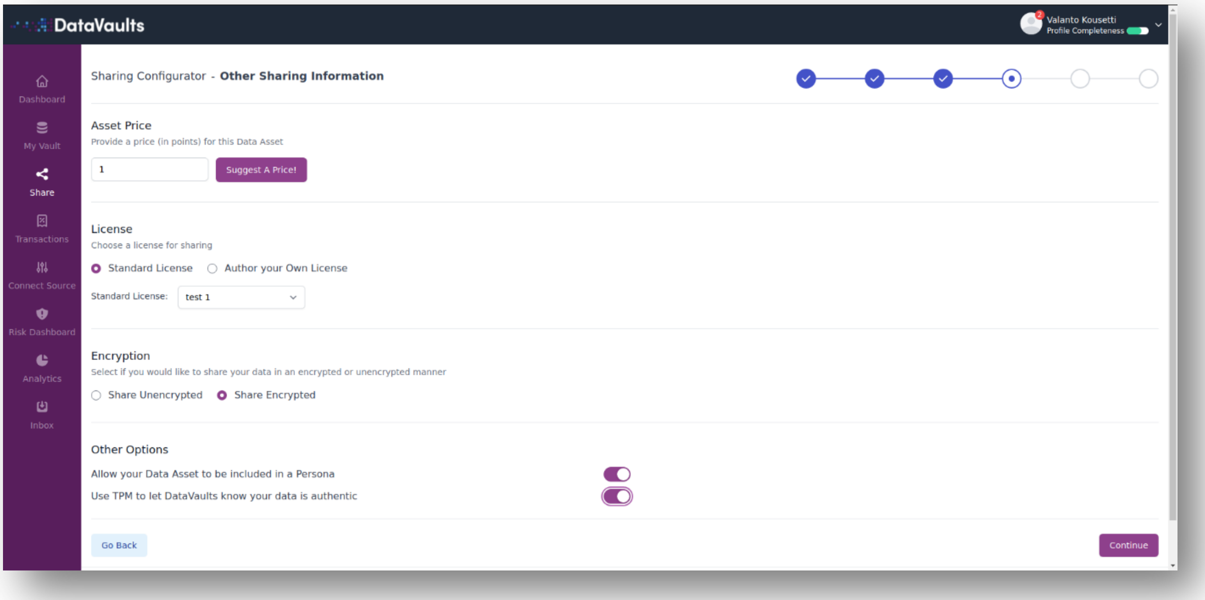
Data Preview

NAME	BIRTHDATE	LOCATION
Ada Lovelance	December 10, 1975	England
Grace Hopper	December 9, 1934	USA
Margaret Hamilton	August 17, 1983	Poland

Anonymisation Levels

Figure 3: Sharing Configurator - Anonymisation Selection

The access policies for this specific data set are provided and finally details on pricing, license, etc. are defined before sharing (Figure 4).



Sharing Configurator - **Other Sharing Information**

Asset Price

Provide a price (in points) for this Data Asset

1

License

Choose a license for sharing

● Standard License ○ Author your Own License

Standard License: test 1

Encryption

Select if you would like to share your data in an encrypted or unencrypted manner

○ Share Unencrypted ● Share Encrypted

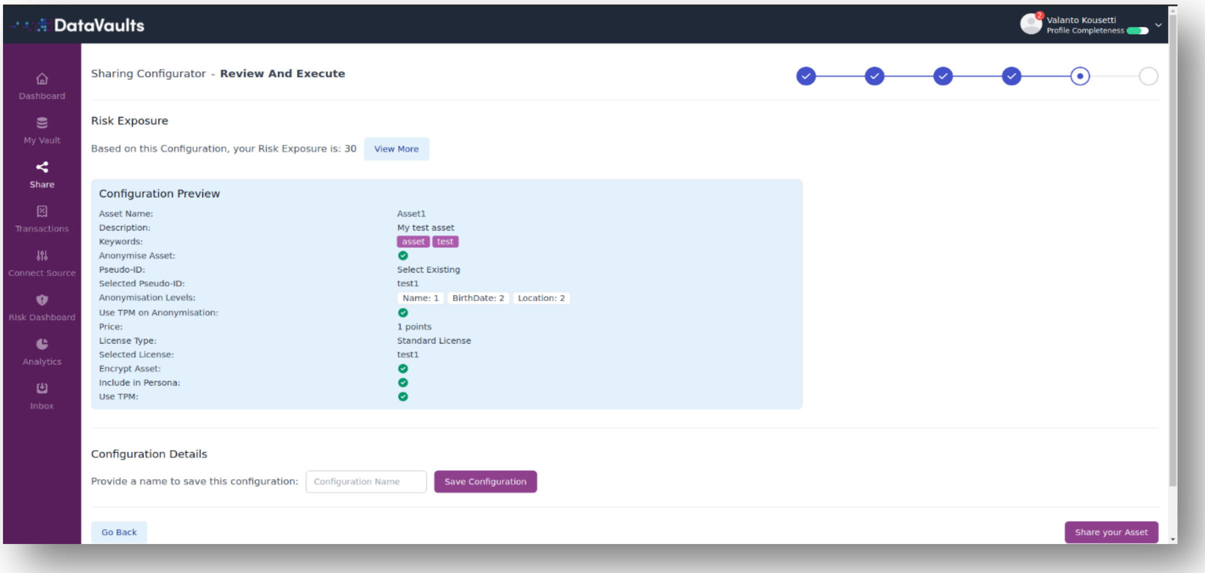
Other Options

Allow your Data Asset to be included in a Persona ●

Use TPM to let DataVaults know your data is authentic ●

Figure 4: Sharing Configurator - Other Sharing Information

Finally, a summary of the selected configuration is displayed before the user can confirm the sharing operation.



Sharing Configurator - **Review And Execute**

Risk Exposure

Based on this Configuration, your Risk Exposure is: 30

Configuration Preview

Asset Name:	Asset1
Description:	My test asset
Keywords:	asset test
Anonymise Asset:	✓
Pseudo-ID:	Select Existing
Selected Pseudo-ID:	test1
Anonymisation Levels:	Name: 1 BirthDate: 2 Location: 2
Use TPM on Anonymisation:	✓
Price:	1 points
License Type:	Standard License
Selected License:	test1
Encrypt Asset:	✓
Include in Persona:	✓
Use TPM:	✓

Configuration Details

Provide a name to save this configuration: Configuration Name

Figure 5: Sharing Configurator – Configuration Preview

2.2.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: Sharing Configurator is a component that is built with the VueJS 3 framework and is using information coming from other components, the Personal App's MongoDB and the Personal App's Postgres database.

2.2.2 Component Backlog

2.2.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project. New user stories that were not part of the initial backlog provided in D3.1 are marked with an asterisk (*) next to their ID.

ID#	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_046	Sharing Configurator. Sharing Setup Manager	Data Provider	create a new sharing configuration, with all fields empty	I can share/upload to the DataVaults Cloud a data asset
US_047	Sharing Configurator. Sharing Setup Manager	Data Provider	select inside the sharing configuration the data asset to be shared	I can make a copy of the selected data asset to the DataVaults Cloud for private or sharing purposes.
US_048	Sharing Configurator. Sharing Setup Manager	Data Provider	select inside the sharing configuration the data source to be shared	I can make copies of data collected from the selected data source to the DataVaults Cloud for private or sharing purposes.
US_049	Sharing Configurator. Sharing Setup Manager	Data Provider	select a data asset to be uploaded and/or shared to the DataVaults Cloud, from my Personal DataVaults App data assets	I can make a copy of the selected data asset to the DataVaults Cloud for private or sharing purposes.
US_050	Sharing Configurator. Sharing Setup Manager	Data Provider	select a connected data source to share data assets collected from there	I can make copies of data collected from this source to the DataVaults Cloud for private or sharing purposes.
US_051	Sharing Configurator. Sharing Setup Manager	Data Provider	select the anonymisation level of the selected data asset	I can choose what happens with the personally identifiable information that is available in the data asset that will be shared.
US_052	Sharing Configurator. Sharing Setup Manager	Data Provider	select the sharing level for the selected data asset (public, private - access policies in effect, confidential -not to be shared)	I can choose whether the specific data asset will be available to others or will only be uploaded for me.

US_053	Sharing Configurator. Sharing Setup Manager	Data Provider	select a time period for which the data asset will be available on the DataVaults Cloud for sharing	I control the availability of my data.
US_054	Sharing Configurator. Sharing Setup Manager	Data Provider	select what part of the data asset and its metadata will be publicly available for preview purposes	I control what is available from my data and at the same time higher the chances of my data being purchased by Data Seekers.
US_055	Sharing Configurator. Sharing Setup Manager	Data Provider	select from a list of predefined licences that should be in effect for the shared data, regarding usage (e.g., no distribution, sharing with reference to source etc.), time (e.g., one-month, one year, forever)	I control that my data assets that are shared are used properly.
US_057	Sharing Configurator. Sharing Setup Manager	Data Provider	view a suggested price for my data asset under the selected sharing options	I can decide easier on a price that maximises my earnings, while remaining competitive in the personal data market ecosystem.
US_058	Sharing Configurator. Sharing Setup Manager	Data Provider	select a price tag for the data asset, under the selected sharing configuration (licence, anonymisation level etc.)	I can be compensated whenever the data asset is acquired by a Data Seeker.
US_060	Sharing Configurator. Sharing Setup Manager	Data Provider	execute the configured sharing	my data asset is uploaded to the DataVaults Cloud Platform.
US_067	Sharing Configurator. Sharing Setup Manager	Data Provider	load a saved sharing configuration for a new data asset to be shared	I can reuse an existing configuration and make any adaptations needed for the new data asset to be shared.
US_072	Access Policy Editor	DataVaults Personal App	Receive the identification of the Individual and load the access policies on the Access Policy Editor interface	The Individual can configure the policies for granting access to her data.
US_073	Access Policy Editor	Data Provider	edit the access policies that apply to my data assets	I change the terms for providing access to my data
US_074	Access Policy Editor	Data Provider	load existing access policy templates for creating new policies	I can easily define the access policies that will apply to my data.
US_075	Access Policy Editor	Data Provider	create new templates from my policies	I can re-use it in the future.
US_076	Access Policy Editor	Data Provider	finalise the policies configuration of a data sharing configuration.	these policies take effect once the sharing configuration is executed.

US_066	Sharing Configurator. Sharing Setup Manager	Data Provider	delete a sharing configuration	I discard any sharing configuration I no longer wish to use.
US_068	Attestation	Data Provider	Authenticate my device in a privacy-preserving way	my data cannot be linked directly to my identity.
US_061	Sharing Configurator. Sharing Setup Manager	Data Provider	Modify the sharing parameters of a data asset I have already shared	my data asset is from this point shared under the new terms.
US_062	Sharing Configurator. Sharing Setup Manager	Data Provider	save the sharing configuration at any stage	I can return to it at a later stage to continue or reload it for reuse.
US_071	Attestation	Data Provider	have the strictest privacy-preserving technology enabled by default	I have the best privacy preservation without further configuration.
US_059	Sharing Configurator. Sharing Setup Manager	Data Provider	Select if I want DAA to be used during data sharing	my data asset will have an extra “trustworthiness” guarantee.
US_251*	Sharing Configurator. Sharing Setup Manager	Data Provider	Be able to use a template with ready-made configurations	I can accelerate the sharing of my data

2.2.2.2 *Features planned for upcoming Releases*

The table below provides the list of features that remain in the backlog and are scheduled to be delivered in the next period, by the integration work to be performed in WP5.

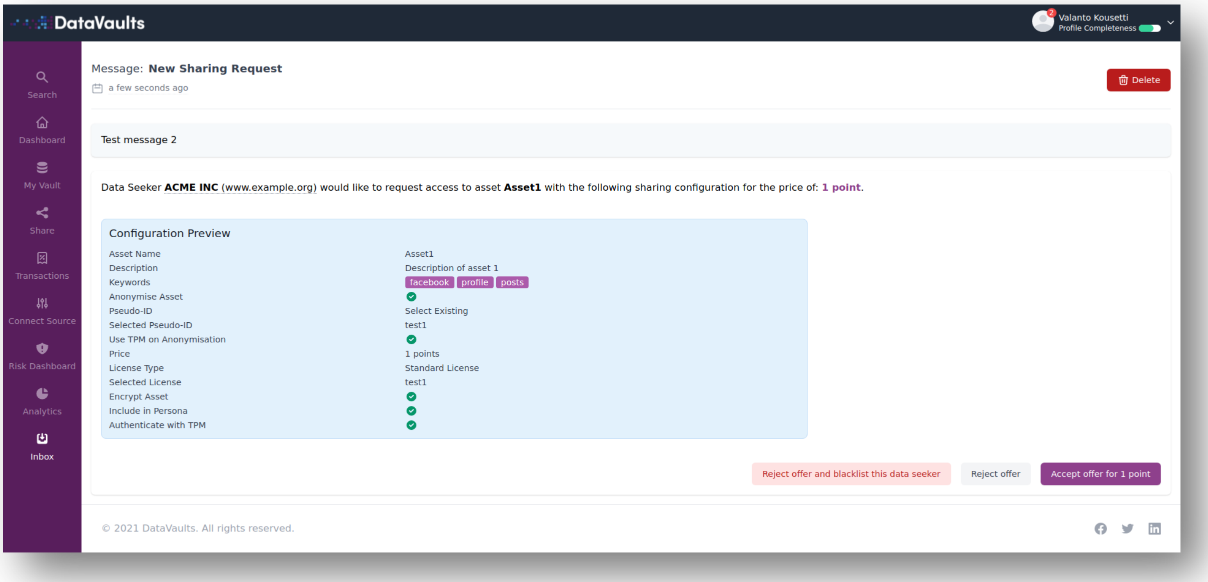
ID#	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_057	Sharing Configurator. Sharing Setup Manager	Data Provider	view a suggested price for my data asset under the selected sharing options	I can decide easier on a price that maximises my earnings, while remaining competitive in the personal data market ecosystem.

2.3 DATA REQUEST SERVICE RESOLVER

Component's Concept Update from D3.2: None

Component Description: The Data Request Service Resolver is the component responsible for getting a data sharing request coming from a Data Seeker and translating this to a sharing configuration proposition to display to the Data Owner, who has the final say regarding accepting or rejecting such a contract.

As such, the data request is presented to the Data Owner, alongside with a message from the Data Seeker as shown in the next figures. The relevant information is displayed in a clear way, to allow users of the target audience to identify the requested asset and understand the implications of the sharing activity. They have the option to accept this request or reject it.



Message: **New Sharing Request**

a few seconds ago

Test message 2

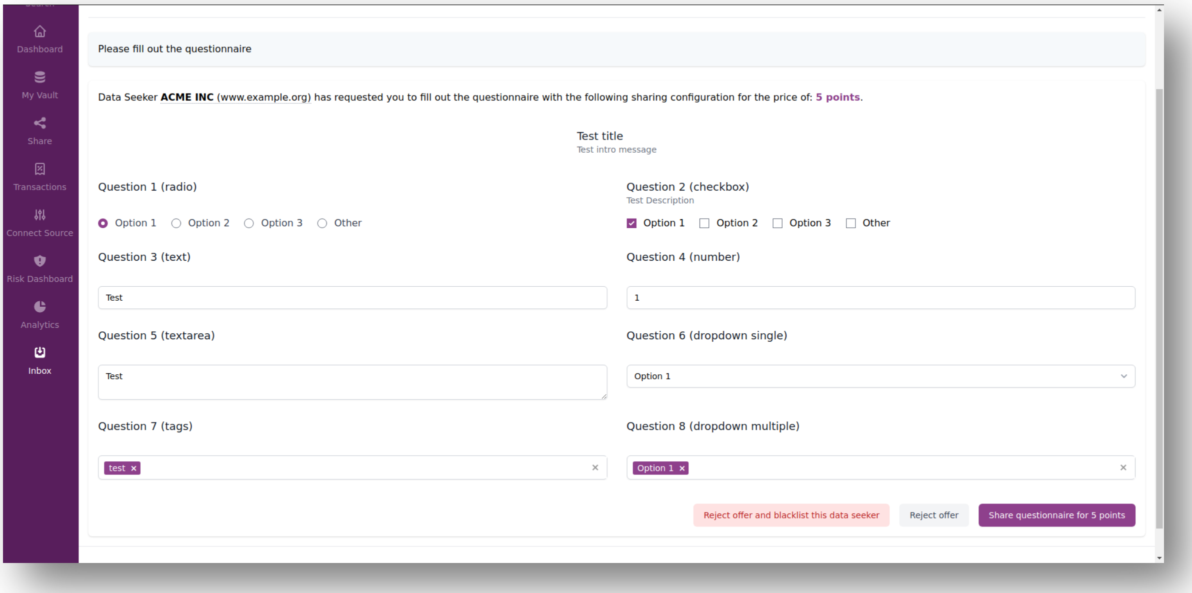
Data Seeker **ACME INC** (www.example.org) would like to request access to asset **Asset1** with the following sharing configuration for the price of: **1 point**.

Configuration Preview

Asset Name	Asset1
Description	Description of asset 1
Keywords	facebook profile posts
Anonymise Asset	✓
Pseudo-ID	Select Existing
Selected Pseudo-ID	test1
Use TPM on Anonymisation	✓
Price	1 points
License Type	Standard License
Selected License	test1
Encrypt Asset	✓
Include in Persona	✓
Authenticate with TPM	✓

Figure 6: View an open data sharing request

In case the request contains a questionnaire, then this component renders the questionnaire at the side of the user, as illustrated in Figure 7.



Please fill out the questionnaire

Data Seeker **ACME INC** (www.example.org) has requested you to fill out the questionnaire with the following sharing configuration for the price of: **5 points**.

Test title
Test intro message

Question 1 (radio)

● Option 1 ○ Option 2 ○ Option 3 ○ Other

Question 2 (checkbox)
Test Description

☑ Option 1 □ Option 2 □ Option 3 □ Other

Question 3 (text)

Test

Question 4 (number)

1

Question 5 (textarea)

Test

Question 6 (dropdown single)

Option 1

Question 7 (tags)

test ×

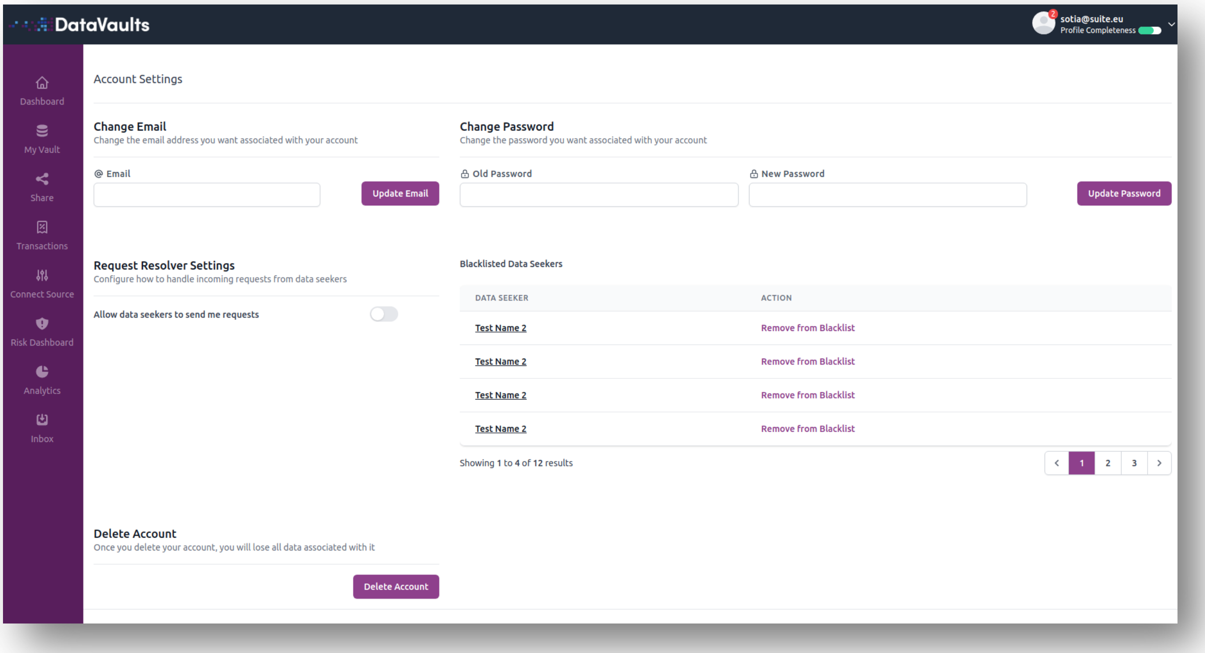
Question 8 (dropdown multiple)

Option 1 ×

Reject offer and blacklist this data seeker
Reject offer
Share questionnaire for 5 points

Figure 7: View of an example questionnaire

To mitigate the occurrence of unwanted data requests, the Personal App can be configured to automatically reject all requests from certain Data Seekers.



Account Settings

Change Email
Change the email address you want associated with your account

@ Email

Change Password
Change the password you want associated with your account

Old Password

New Password

Request Resolver Settings
Configure how to handle incoming requests from data seekers

Allow data seekers to send me requests

Blacklisted Data Seekers

DATA SEEKER	ACTION
Test Name 2	Remove from Blacklist
Test Name 2	Remove from Blacklist
Test Name 2	Remove from Blacklist
Test Name 2	Remove from Blacklist

Showing 1 to 4 of 12 results

Delete Account
Once you delete your account, you will lose all data associated with it

Figure 8: Blacklisting Data Seekers and Service Toggling

2.3.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The Data Request Service Resolver is built with the VueJS 3 framework and uses information coming from other components, the Personal App's MongoDB and the Personal App's Postgres database.

The messaging protocol for getting such requests from the cloud-based infrastructure is based on RabbitMQ.

2.3.2 Component Backlog

2.3.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_100	Data Request Service Resolver	Data Provider	receive a notification on my Personal DataVaults App whenever a custom request for my data is made	I am instantly informed of any pending requests.
US_101	Data Request Service Resolver	Data Provider	view the details of the sharing proposal made by the Data Seeker, including the requested type of data, usage, licence, price, seeker's information (organisation etc.).	I have a full overview of the sharing & usage terms prior to accepting or rejecting the proposal.
US_102	Data Request Service Resolver	Data Provider	accept the sharing proposal/request	the sharing of my data asset under the accepted terms can take place.
US_103	Data Request Service Resolver	Data Provider	reject the sharing proposal/request	I can keep having some data only on the personal DataVaults side and not share them with anyone else.

US_104	Data Request Service Resolver	Data Provider	modify the notification settings	I can block recurrent requests without disabling the service completely.
US_105	Data Request Service Resolver	Data Provider	disable the service completely	I will not get any notifications in the future.
US_106	Data Request Service Resolver	Data Provider	blacklist certain Data Seekers or business sectors from performing requests	I quickly filter out Data Seekers with whom I am not interested to share data

2.3.2.2 Features planned for upcoming releases

All identified features of the initial platform requirements have been delivered. If new user stories are identified during development and demonstrator validation (WP6), they will be included in subsequent deliverables of WP5.

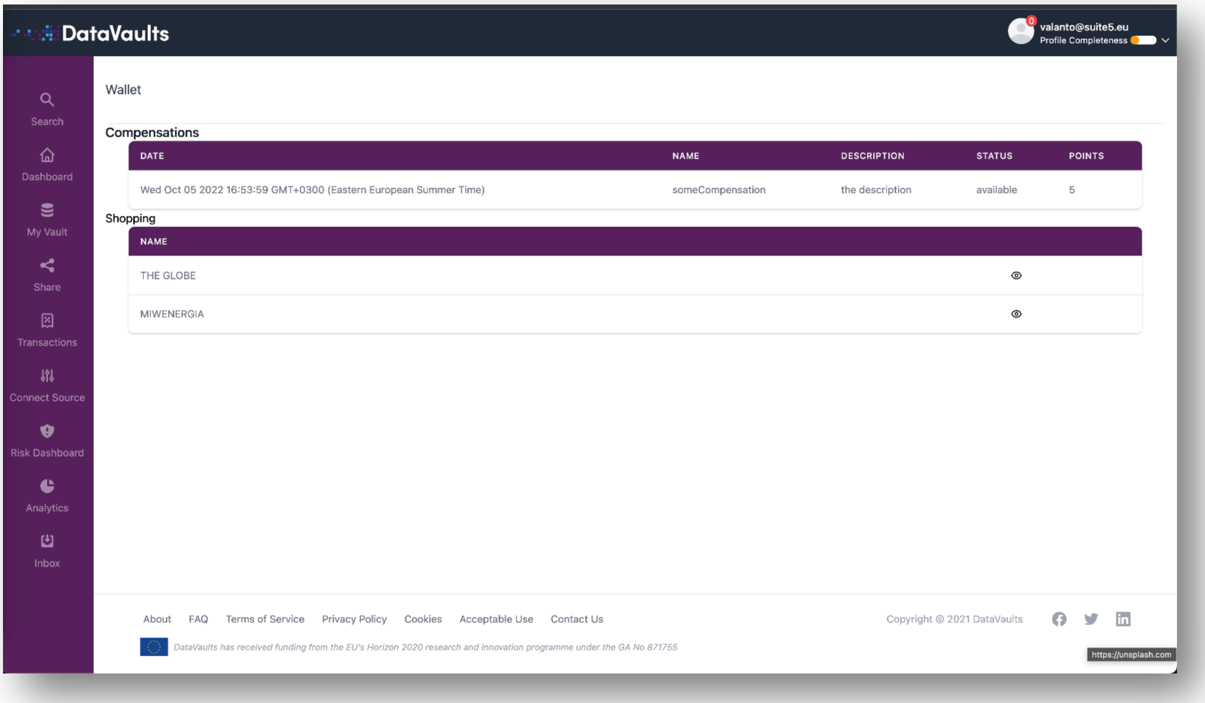
2.4 PERSONAL WALLET

<u>Component's Concept Update from D3.2:</u> None

Component Description: The personal wallet module provides the functionality required by Data Owners to receive and use compensations they receive as result of sharing data while keeping privacy of users during the whole flow of compensations, since they are received and transferred to the wallet until they are used.

The wallet also provides an interface for merchants to offer products to be acquired, or services used with compensations. This is enabled by publishing products and performing the compensation exchange for products or services through the DataVaults backend services.

On the main window of the Wallet component, the user can choose to see the details of completed transactions or go to the merchant pages.



Wallet

Compensations

DATE	NAME	DESCRIPTION	STATUS	POINTS
Wed Oct 05 2022 16:53:59 GMT+0300 (Eastern European Summer Time)	someCompensation	the description	available	5

Shopping

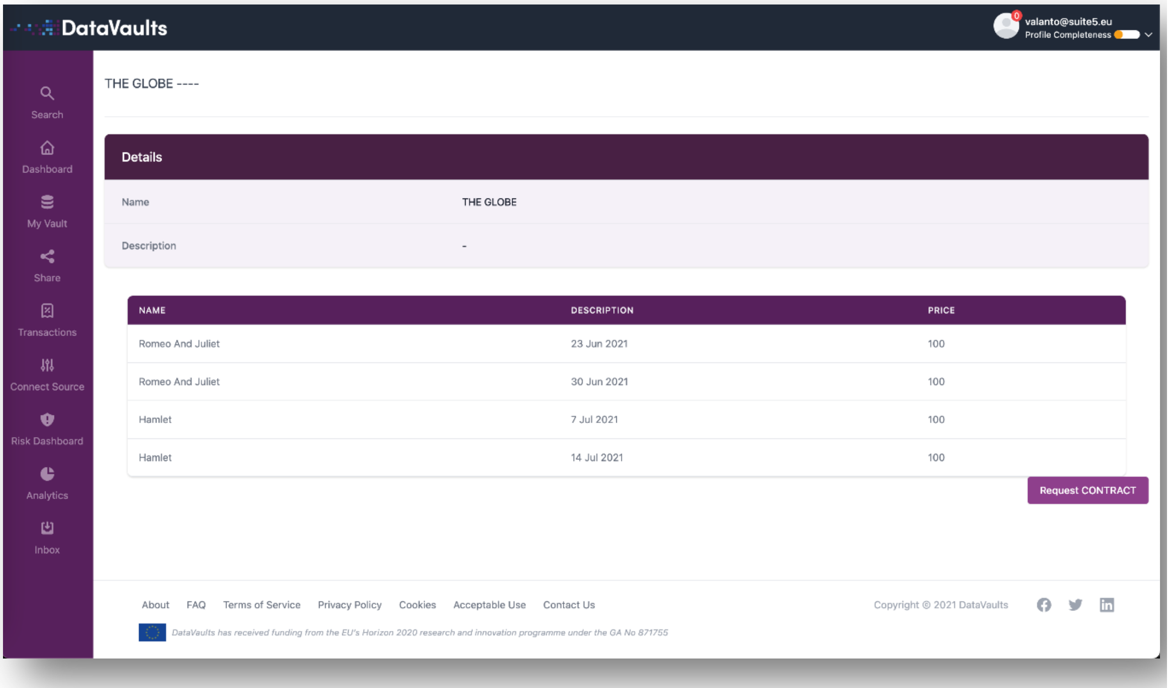
NAME	
THE GLOBE	👁
MIWENERGIA	👁

Copyright © 2021 DataVaults

DataVaults has received funding from the EU's Horizon 2020 research and innovation programme under the GA No 871755

Figure 9: Personal Wallet compensation overview

The Personal Wallet component will enable multiple merchants to offer their goods and services. The user can then exchange previously received compensation for a range of products (Figure 10 and Figure 11).



THE GLOBE ----

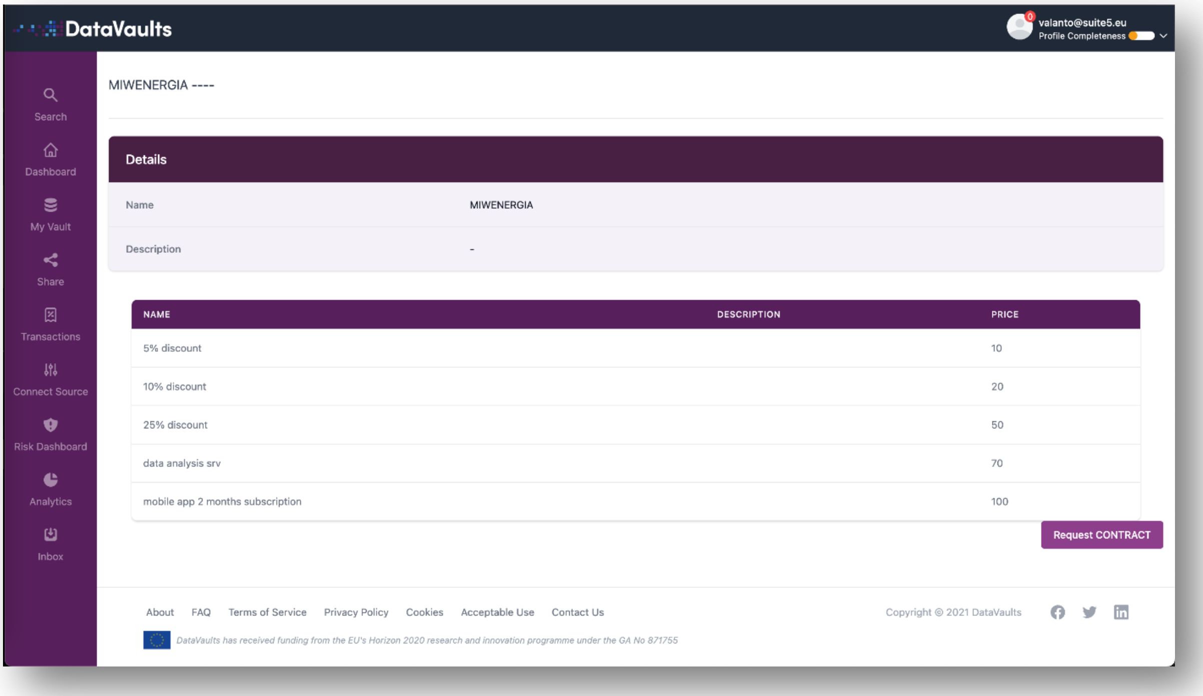
Details

Name	THE GLOBE
Description	-

NAME	DESCRIPTION	PRICE
Romeo And Juliet	23 Jun 2021	100
Romeo And Juliet	30 Jun 2021	100
Hamlet	7 Jul 2021	100
Hamlet	14 Jul 2021	100

DataVaults has received funding from the EU's Horizon 2020 research and innovation programme under the GA No 871755

Figure 10: Personal Wallet merchant purchase



MIWENERGIA ----

Details

Name	MIWENERGIA
Description	-

NAME	DESCRIPTION	PRICE
5% discount		10
10% discount		20
25% discount		50
data analysis srv		70
mobile app 2 months subscription		100

DataVaults has received funding from the EU's Horizon 2020 research and innovation programme under the GA No 871755

Figure 11: Personal Wallet merchant purchase 2

2.4.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The personal wallet is based on the development carried out in the *Functional Encryption TEChnologies* (FENTEC) project for the Privacy Enhanced Digital Currency Prototype[2]. This tool will be integrated with the Blockchain Security2Go Starter Kit to improve the security of the encryption scheme related to the creation of crypto tokens and their usage.

2.4.2 Component Backlog

2.4.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_196	Personal DataVaults Wallet	Data Provider	display the value of all accounts in my wallet	I can evaluate the worth of the previous sharing activities.
US_198	Personal DataVaults Wallet	Data Provider	exchange the funds into real-world goods offered through the platform	a real value is obtained from the data.
US_201	Personal DataVaults Wallet	Data Provider	spend my earnings, preserving my anonymity	the use of compensations does not leak information about me
US_197	Personal DataVaults Wallet	Data Provider	create a new account	a new pseudonym is used for subsequent data uploads.

2.4.2.2 Features planned for upcoming Releases

The table below provides the list of features that remain in the backlog and are scheduled to be delivered in the next period, by the integration work to be performed in WP5.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_194	Personal DataVaults Wallet	Data Provider	use multiple Blockchain addresses	I keep the data from unrelated sources separately.

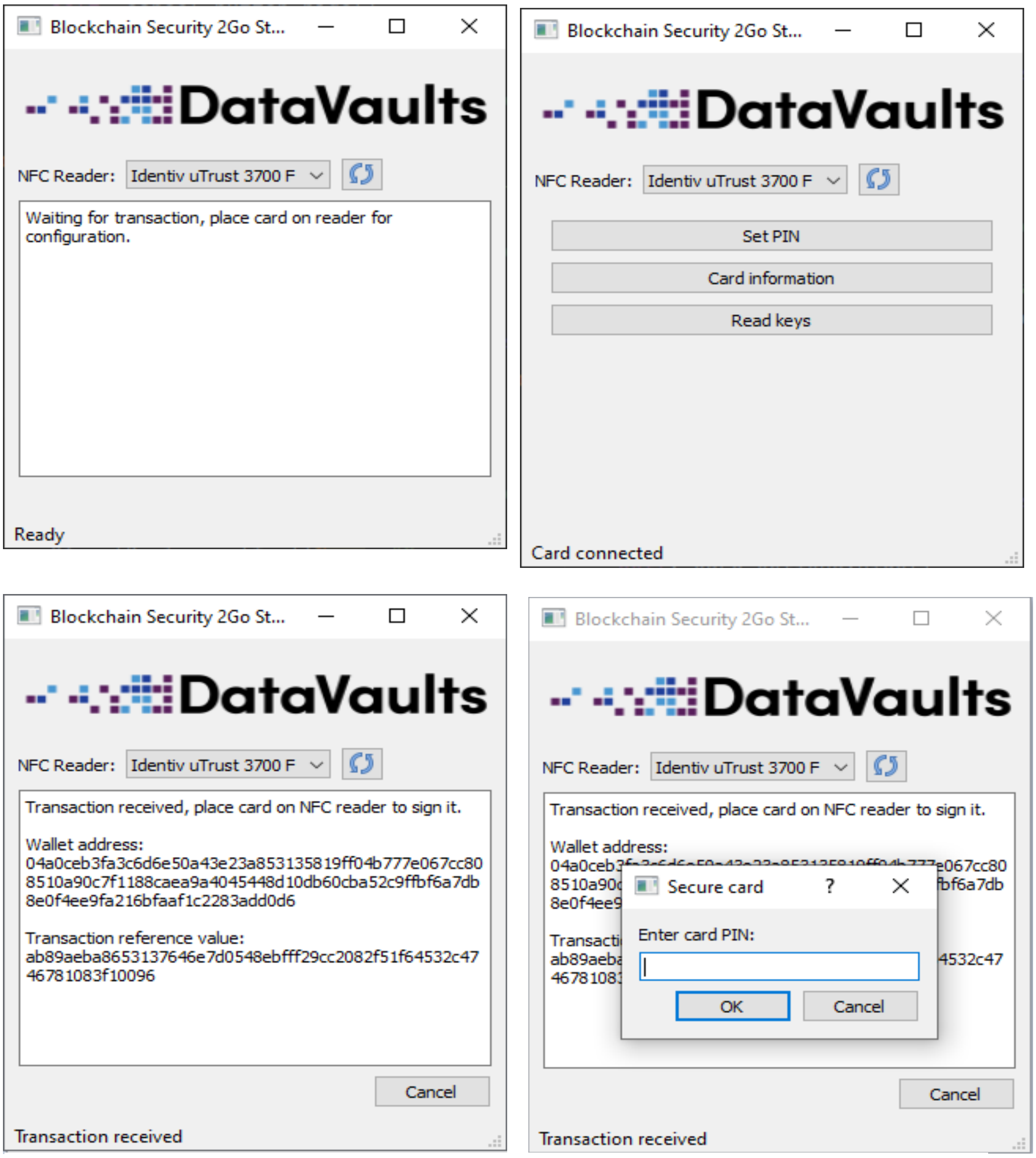
[2] http://fentec.eu/sites/default/files/fentec/public/content-files/deliverables/FENTEC_D7.5_v1.0.pdf

2.5 Blockchain Security 2Go Starter Kit

Component's Concept Update from D3.2: None

Component Description: The Blockchain Security 2Go Starter Kit is a smart card with Near Field Communication (NFC) interface for securely managing keys and creating signatures. It is optionally used to sign the Blockchain transactions from the user, e.g. new sharing configurations. In DataVaults, it is additionally utilized to manage key pairs for the previously described Personal Wallet component.

When a new sharing configuration is created, or a payment is started, a signature request is forwarded to a program installed on the user's device. It will display a notification about the new request, show a control value and ask the user to hold the card to the reader. The signature is then returned to the Personal App, which then forwards it to the cloud backend.



Waiting for transaction, place card on reader for configuration.

Ready

Card connected

Transaction received, place card on NFC reader to sign it.

Wallet address:
04a0ceb3fa3c6d6e50a43e23a853135819ff04b777e067cc80
8510a90c7f1188caea9a4045448d10db60cba52c9ffbf6a7db
8e0f4ee9fa216bfaaf1c2283add0d6

Transaction reference value:
ab89aeba8653137646e7d0548ebfff29cc2082f51f64532c47
46781083f10096

Transaction received

Enter card PIN:

Figure 12: User interface for interacting with the Blockchain Security 2Go Starter Kit

2.5.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: A locally installed program forwards the communication from the web browser to the Blockchain Security 2Go Starter Kit through an NFC-reader. This software, called *Bridge*, listens for requests from the DataVaults Personal App with a local web server and initializes the low-level interface to the NFC reader.

The *Bridge* is written in Python and uses the pyscard and blocksec2go[3] libraries for the communication with the smart card. A Flask server is providing the API for the Personal App, PySide2 is included as GUI framework.

2.5.2 Component Backlog

2.5.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_195	Personal DataVaults Wallet	Data Provider	see a list of all my blockchain addresses	I can view my previously used addresses.
US_199	Personal DataVaults Wallet	Data Provider	securely store the private keys for my accounts	they cannot be leaked by other applications on the device.
US_200	Notification System	Data Provider	have a clear indication when my private key is used for signing a Blockchain transaction	no immutable actions are performed accidentally.

2.5.2.2 Features planned for upcoming Releases

All identified features of the initial platform requirements have been delivered. If new user stories are identified during development and demonstrator validation (WP6), they will be included in subsequent deliverables.

2.5.2.3 Optional extensions with biometric authentication

For further improvements in the usability and privacy of the architecture, we propose biometric user-authentication incorporated on the smart card hosting the private keys for signatures needed to initiate Blockchain or smart contract transactions. If the contactless

[3] https://github.com/Infineon/BlockchainSecurity2Go-Python-Library

smartcard as provided in the Security2Go Starter Kit stays the only single factor for authentication, lost or stolen cards imply a high risk for abuse of those cards (if no PIN authentication is enforced). Thus, multi-factor authentication is required to increase security. Passwords as one of more factors have shown high security risks due to simplification, i.e. short passwords and passwords that can be easily found in dictionaries. Passphrases – the usually somewhat more secure version of passwords – are still error-prone, maybe forgotten or stolen if written on paper. Both, passwords and passphrases have provided terrible user-experience when needed to be typed in on small screen keyboards like on mobile devices.
Biometrics seem to be the solution, but often the implementation of biometrics as second factor need an either untrusted or expensive terminal with the fingerprint sensor and the appropriate feature extraction algorithms.

With biometrics on card, many of these problems can be solved. Smart cards with integrated biometric sensors evaluate the fingerprint inside the secured card and initiate a cryptographically strong authentication protocol after successful fingerprint match.

Figure 13: Fingerprint card as privacy preserving authentication token

In contrast to other applications like biometric payment, the focus in DataVaults can be twofold: On the one hand to provide biometric two-factor authentication for signature generation of Blockchain/smart contract transactions, and on the other hand to initiate a privacy-preserving anonymous or pseudonymous attestation to allow for the untraceable verification of legitimate users of the DataVaults platform.

2.6 DATA ANONYMIZER

Component's Concept Update from D3.2: None

Component Description: The anonymizer is the component of DataVaults responsible for preserving data privacy. It alters the data in such a way, that it will preserve its usefulness but hide the original data. With these modifications, it cannot be traced back to the individuals the data was taken from.

The anonymizer is capable of taking a dataset and obfuscating the contained data by replacing it with values that represent the original data in a way that is non-identifying (e.g. an age of 29 may be replaced with [20-30] or a name Darren Smith may be replaced with Darren *****).

Via the frontend interface, users will be able to configure the anonymization process by selecting different anonymization pre-sets, which can be applied to a column in their dataset or they can use advanced settings to allow for more configurability in their anonymization. To see how their choices will impact the final result, a preview button is available. Upon clicking this button, users will see a subset of their dataset with their current anonymization options applied to it. This will help users understand the impact of their choices.

The PseudoID generator is a smaller component, capable of producing a unique ID for a user who wishes to share their data as an anonymous user. This ID may then be used for the purposes of communication with the data owner whilst preserving their anonymity.

In this version of the Anonymiser management of 'Location Privacy' is also supported. 'Location Privacy' is defined as "the ability of an individual to move in public space with the expectation that under normal circumstances their location will not be systematically and secretly recorded for later use"[4]

The existence of location databases stripped of identifying tags can leak information. For instance, if I know that Vera is the only person who lives on Dead End Lane, the datum that someone used a location-based service on Dead End Lane can be reasonably linked to Vera.

Since location privacy definition and requirements differ depending on the scenario, no single technique is able to address the requirements of all location privacy categories. Therefore, in the past, the research community, focusing on providing solutions for the protection of location privacy of users, has defined techniques that can be divided into three main classes: *anonymity-based, obfuscation-based*, and *policy-based* techniques. These classes of techniques are partially overlapped in scope and could be potentially suitable to cover requirements coming from one or more of the categories of location privacy.

It is easy to see that anonymity-based and obfuscation-based techniques can be considered dual categories. Anonymity-based techniques have been primarily defined to protect identity privacy and are not suitable for protecting position privacy, whereas obfuscation-based

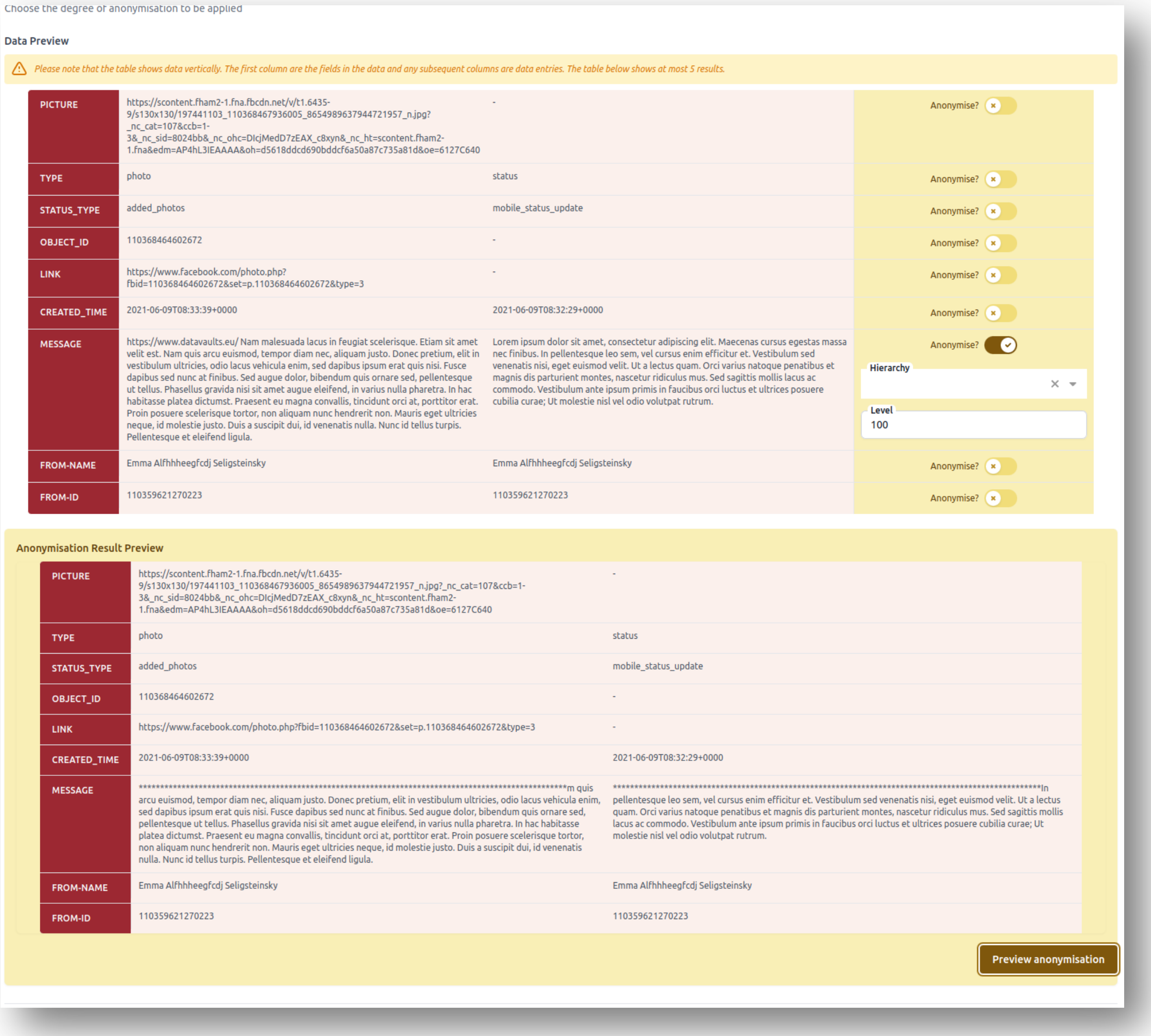
[4] https://www.eff.org/wp/locational-privacy

techniques are well suited for position protection and not appropriate for identity protection. Anonymity-based and obfuscation-based techniques could also be both exploited for protecting path privacy. Policy-based techniques are in general suitable for all location privacy categories, although they are often difficult to understand and manage for end users.

Within the context of DataVaults, it is important to consider the notion of utility within the context of anonymising location data – if the data seeker is looking to understand different groups mobility patterns to inform public transport planning for example accurate location data over time at scale is imperative. Therefore, supporting location privacy has to also consider the impact on the utility of that data – it will impact the monetary value of that data if the necessary insights can no longer be reliably derived from that data.

Location Privacy in DataVaults as a result is supported through a mechanism to generate new **Synthetic Data** that has the **same format and statistical properties** as the original location data.

Synthetic data can then be used to supplement, augment and in some cases replace real data when training Machine Learning models. Additionally, it enables the testing of Machine Learning or other data dependent software systems without the risk of exposure that comes with data disclosure.



Choose the degree of anonymisation to be applied

Data Preview

⚠ *Please note that the table shows data vertically. The first column are the fields in the data and any subsequent columns are data entries. The table below shows at most 5 results.*

PICTURE	https://scontent.fham2-1.fna.fbcdn.net/v/t1.6435-9/s130x130/197441103_110368467936005_8654989637944721957_n.jpg?_nc_cat=107&ccb=1-3&_nc_sid=8024bb&_nc_ohc=DIcjMedD7zEAX_c8xyn&_nc_ht=scontent.fham2-1.fna&edm=AP4hL3IEAAAA&oh=d5618ddcd690bddcf6a50a87c735a81d&oe=6127C640	-	Anonymise? ✕
TYPE	photo	status	Anonymise? ✕
STATUS_TYPE	added_photos	mobile_status_update	Anonymise? ✕
OBJECT_ID	110368464602672	-	Anonymise? ✕
LINK	https://www.facebook.com/photo.php?fbid=110368464602672&set=p.110368464602672&type=3	-	Anonymise? ✕
CREATED_TIME	2021-06-09T08:33:39+0000	2021-06-09T08:32:29+0000	Anonymise? ✕
MESSAGE	https://www.datavaults.eu/ Nam malesuada lacus in feugiat scelerisque. Etiam sit amet velit est. Nam quis arcu euismod, tempor diam nec, aliquam justo. Donec pretium, elit in vestibulum ultricies, odio lacus vehicula enim, sed dapibus ipsum erat quis nisi. Fusce dapibus sed nunc at finibus. Sed augue dolor, bibendum quis ornare sed, pellentesque ut tellus. Phasellus gravida nisi sit amet augue eleifend, in varius nulla pharetra. In hac habitasse platea dictumst. Praesent eu magna convallis, tincidunt orci at, porttitor erat. Proin posuere scelerisque tortor, non aliquam nunc hendrerit non. Mauris eget ultricies neque, id molestie justo. Duis a suscipit dui, id venenatis nulla. Nunc id tellus turpis. Pellentesque et eleifend ligula.	Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas cursus egestas massa nec finibus. In pellentesque leo sem, vel cursus enim efficitur et. Vestibulum sed venenatis nisi, eget euismod velit. Ut a lectus quam. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed sagittis mollis lacus ac commodo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Ut molestie nisl vel odio volutpat rutrum.	Anonymise? ✓ Hierarchy ✕ Level 100
FROM-NAME	Emma Alfhhheegfcdj Seligsteinsky	Emma Alfhhheegfcdj Seligsteinsky	Anonymise? ✕
FROM-ID	110359621270223	110359621270223	Anonymise? ✕

Anonymisation Result Preview

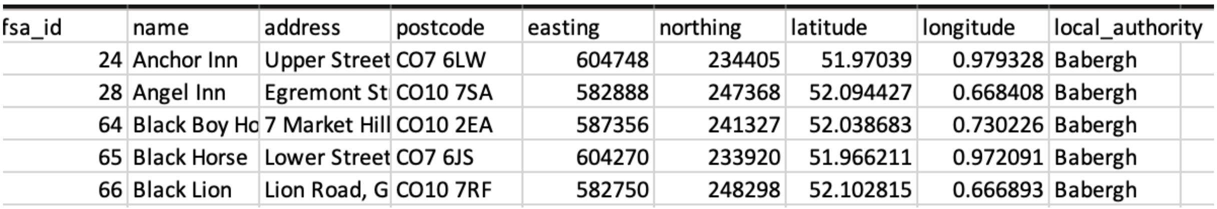
PICTURE	https://scontent.fham2-1.fna.fbcdn.net/v/t1.6435-9/s130x130/197441103_110368467936005_8654989637944721957_n.jpg?_nc_cat=107&ccb=1-3&_nc_sid=8024bb&_nc_ohc=DIcjMedD7zEAX_c8xyn&_nc_ht=scontent.fham2-1.fna&edm=AP4hL3IEAAAA&oh=d5618ddcd690bddcf6a50a87c735a81d&oe=6127C640	-
TYPE	photo	status
STATUS_TYPE	added_photos	mobile_status_update
OBJECT_ID	110368464602672	-
LINK	https://www.facebook.com/photo.php?fbid=110368464602672&set=p.110368464602672&type=3	-
CREATED_TIME	2021-06-09T08:33:39+0000	2021-06-09T08:32:29+0000
MESSAGE	***m quis arcu euismod, tempor diam nec, aliquam justo. Donec pretium, elit in vestibulum ultricies, odio lacus vehicula enim, sed dapibus ipsum erat quis nisi. Fusce dapibus sed nunc at finibus. Sed augue dolor, bibendum quis ornare sed, pellentesque ut tellus. Phasellus gravida nisi sit amet augue eleifend, in varius nulla pharetra. In hac habitasse platea dictumst. Praesent eu magna convallis, tincidunt orci at, porttitor erat. Proin posuere scelerisque tortor, non aliquam nunc hendrerit non. Mauris eget ultricies neque, id molestie justo. Duis a suscipit dui, id venenatis nulla. Nunc id tellus turpis. Pellentesque et eleifend ligula.	**In pellentesque leo sem, vel cursus enim efficitur et. Vestibulum sed venenatis nisi, eget euismod velit. Ut a lectus quam. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed sagittis mollis lacus ac commodo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Ut molestie nisl vel odio volutpat rutrum.
FROM-NAME	Emma Alfhhheegfcdj Seligsteinsky	Emma Alfhhheegfcdj Seligsteinsky
FROM-ID	110359621270223	110359621270223

Figure 14: Anonymizer - Main View

2.6.1 Technology Background

Technology Background Update from D3.2:

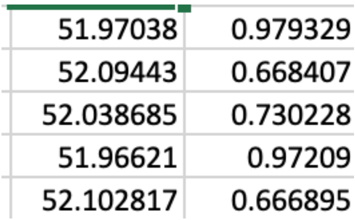
For location privacy, we use probabilistic graphical modelling and deep learning based techniques. To enable a variety of data storage structures, we employ unique hierarchical generative modelling and recursive sampling techniques. We are in essence, learning from real data and generating synthetic clones with high fidelity



fsa_id	name	address	postcode	easting	northing	latitude	longitude	local_authority
24	Anchor Inn	Upper Street	CO7 6LW	604748	234405	51.97039	0.979328	Babergh
28	Angel Inn	Egremont St	CO10 7SA	582888	247368	52.094427	0.668408	Babergh
64	Black Boy Ho	7 Market Hill	CO10 2EA	587356	241327	52.038683	0.730226	Babergh
65	Black Horse	Lower Street	CO7 6JS	604270	233920	51.966211	0.972091	Babergh
66	Black Lion	Lion Road, G	CO10 7RF	582750	248298	52.102815	0.666893	Babergh

Figure 15 Location Dataset Sample

The table above provides a snapshot of a dataset for the location of pubs and bars in the United Kingdom.



51.97038	0.979329
52.09443	0.668407
52.038685	0.730228
51.96621	0.97209
52.102817	0.666895

Figure 16 Synthetically Generated Lat/Long

Above are the latitude and longitude values that are generated through the anonymisation mechanisms developed and integrated into DataVaults – the same statistical properties and fidelity/utility is maintained.

Technology Description: Currently, the anonymizer takes the form of a Java application with 3 classes: AnonHandling contains methods responsible for configuring and executing the anonymization process; BuilderTableHandling contains methods responsible for finding and returning the HierarchyBuilders to the AnonHandling methods, so that they can be used when configuring the anonymization process; and HierBuilding contains methods responsible for generating new HierarchyBuilders, although these methods are currently unused, as these will only become relevant once users are allowed to generate their own HierarchyBuilders.

The final desired workflow for the anonymizer application is documented in the following UML diagram:

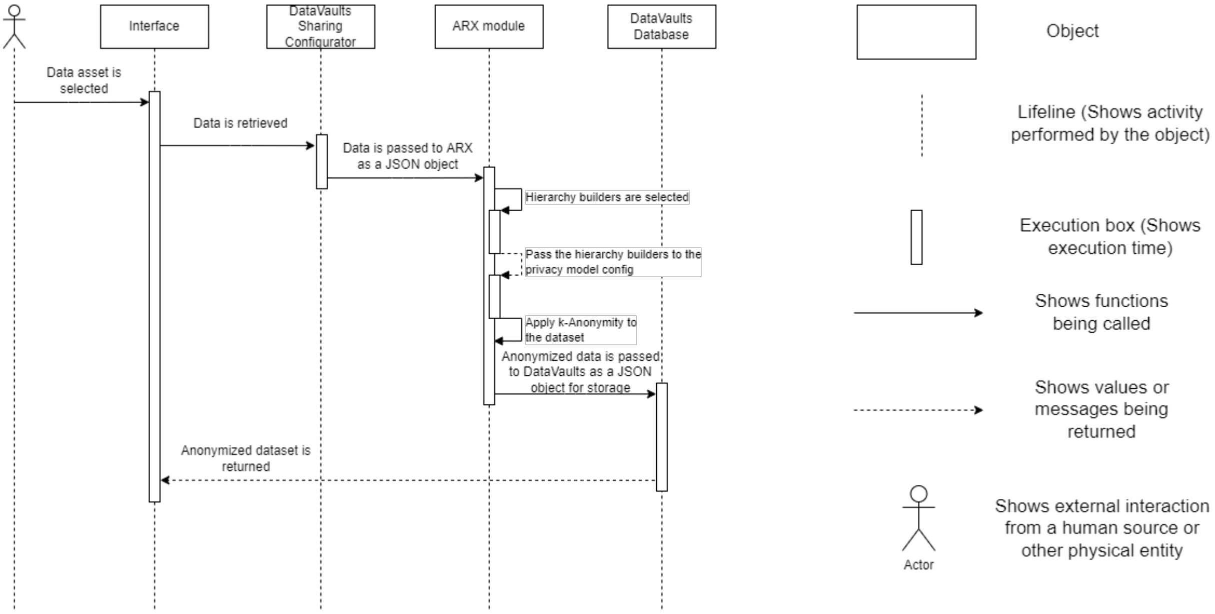

Interface
DataVaults Sharing Configurator
ARX module
DataVaults Database
Data asset is selected
Data is retrieved
Data is passed to ARX as a JSON object
Hierarchy builders are selected
Pass the hierarchy builders to the privacy model config
Apply k-Anonymity to the dataset
Anonymized data is passed to DataVaults as a JSON object for storage
Anonymized dataset is returned
Object
Lifeline (Shows activity performed by the object)
Execution box (Shows execution time)
Shows functions being called
Shows values or messages being returned
Actor
Shows external interaction from a human source or other physical entity

Figure 17: Workflow of Data Anonymizer

Using the hierarchy builder enables the creation of hierarchies on the fly as opposed to utilising pre-defined generated hierarchies that could prove restrictive if new data sources are not supported by the library of hierarchies available. The hierarchy builder will enable the anonymizer to create hierarchies of the following types:

- Redaction based hierarchies
- Interval based hierarchies
- Order based hierarchies
- Date based hierarchies

These take the form of a locally stored *ahs* file that stores information containing the rules required to build a hierarchy during runtime. By importing these hierarchy builders when we need them, we avoid the need to keep hierarchies stored and update them whenever we encounter new values.

2.6.2 Component Backlog

2.6.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_225	Data Anonymiser	Data Seeker	retrieve through an API the list of all data assets available in the	I can utilise the list in another system

			DataVaults Cloud Platform	
US_226	Data Anonymiser	Data Provider	apply anonymisation to my data	personal information can be hidden
US_227	Data Anonymiser	Data Provider	create a fake ID	to hide my real ID that my data belongs to
US_228	Data Anonymiser	Data Provider	be able to select a pre-set anonymisation level	I can anonymise data using a specific approach
US_229	Data Anonymiser	Data Provider	be able to configure the anonymisation approach used on my data	I have flexibility over how my data is shared
US_230	Data Anonymiser	Data Provider	Be able to have more granular control over the anonymisation configuration	I have flexibility over how my data is shared

2.6.2.2 Features planned for upcoming Releases

All identified features of the initial platform requirements have been delivered. If new user stories are identified during development and demonstrator validation (WP6), they will be included in subsequent deliverables.

2.7 Attribute Based Encryption Engine

Component's Concept Update from D3.2: None

Component Description: The ABE engine is one of the modules in DataVaults, which implements sharing data preferences of data owners by providing access management to encrypted data.

This component provides data owners to set access policies to data sets or documents as a whole or by parts. This enables the application of different encryption patterns to data. As it is defined in D2.3, the ABE engine uses patterns to describe the encryption behaviour, to enable users to encrypt a document as a whole with a single access policy or encrypt the same document splitting it in pieces and applying a different access policy to each.

For the current version of the engine, a graphical interface has not been yet defined. Therefore in this first release the ABE engine will use a set of predefined policies to validate the encryption by patterns approach. It will be considered to use the same policy as the one defined in the Policy Editor module. Translation of this policy to the format used in the ABE engine is a requirement, as this is a Boolean expression format which entails more expressiveness limitations.

Key management will be simplified to one Master and public key pair for all participants; therefore, there is no need to implement a secure key hosting service for ABE at this stage.

2.7.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The ABE engine is formed by four components implemented in JAVA and based on the ABE encryption schemes developed in the *Functional Encryption TEChnologies* (FENTEC)[5] project. These components are aimed to cope with each of the next functionalities: Key management, policy management, encryption and decryption.

2.7.2 Component Backlog

2.7.2.1 Implemented Features (delivered in the v0.50 Release)

There have been no updates with respect to information provided in previous deliverables. The table below provides the list of the features that have been delivered as part of the Beta release, as well as the list of features that have been part of the Alpha release. The latter have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

[5] https://fentec.eu/

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_091	ABE Engine	Data Provider	I want to decide who can access my encrypted data	so that I choose who can learn what about me
US_094	ABE Engine	Data Provider	I want to be able to revoke/modify access to my encrypted data	I can re-define access policies applied to my encrypted data
US_098	ABE Engine	Data Seeker	simplify the access to encrypted data as much as possible	so that I do not need to perform extra operations.
US_099	ABE Engine	DataVaults Cloud Platform	I want to be able to decrypt the data using ABE on behalf of a Data Seeker who has purchased it	I provide the data to the Data Seeker

2.7.2.2 Features planned for upcoming Releases

All identified features of the initial platform requirements have been delivered. If new user stories are identified during development and demonstrator validation (WP6), they will be included in subsequent deliverables of WP5.

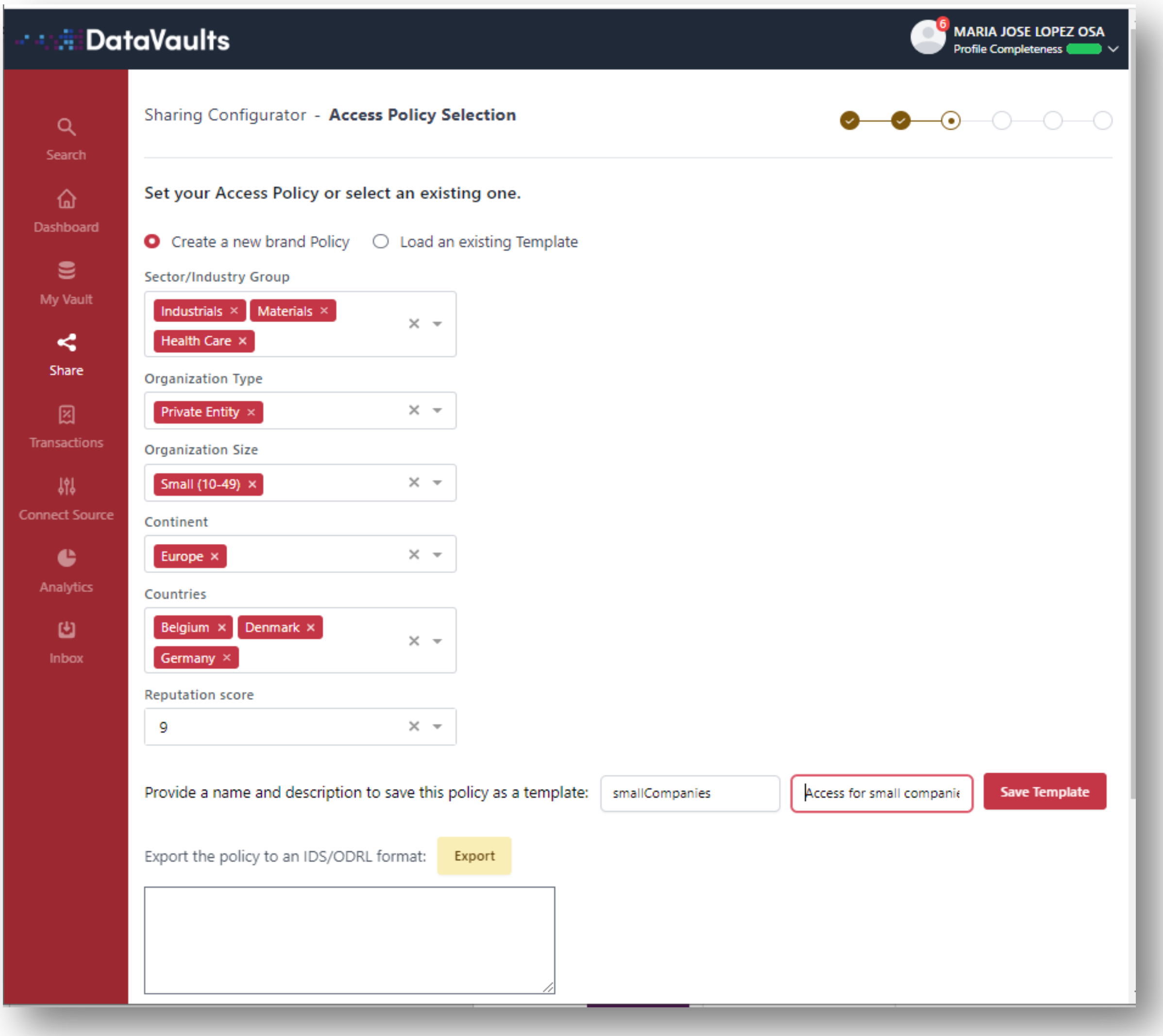
2.8 ACCESS POLICIES EDITOR

Component's Concept Update from D3.2: None

Component Description: The Access Policies Editor is envisioned as part of the Data Sharing configuration in DataVaults. It allows the individuals to define the conditions under which their data will be shared.

The editor shows different attributes for choosing the values that the seekers should meet when accessing the selected Data set. The individual can create a new policy or load one of the locally stored policies for reusing them.

For each attribute the user can select a list of allowed values. These attributes can be seen in Figure 18.



Sharing Configurator - **Access Policy Selection**

Set your Access Policy or select an existing one.

● Create a new brand Policy ○ Load an existing Template

Sector/Industry Group: Industrials ×, Materials ×, Health Care ×

Organization Type: Private Entity ×

Organization Size: Small (10-49) ×

Continent: Europe ×

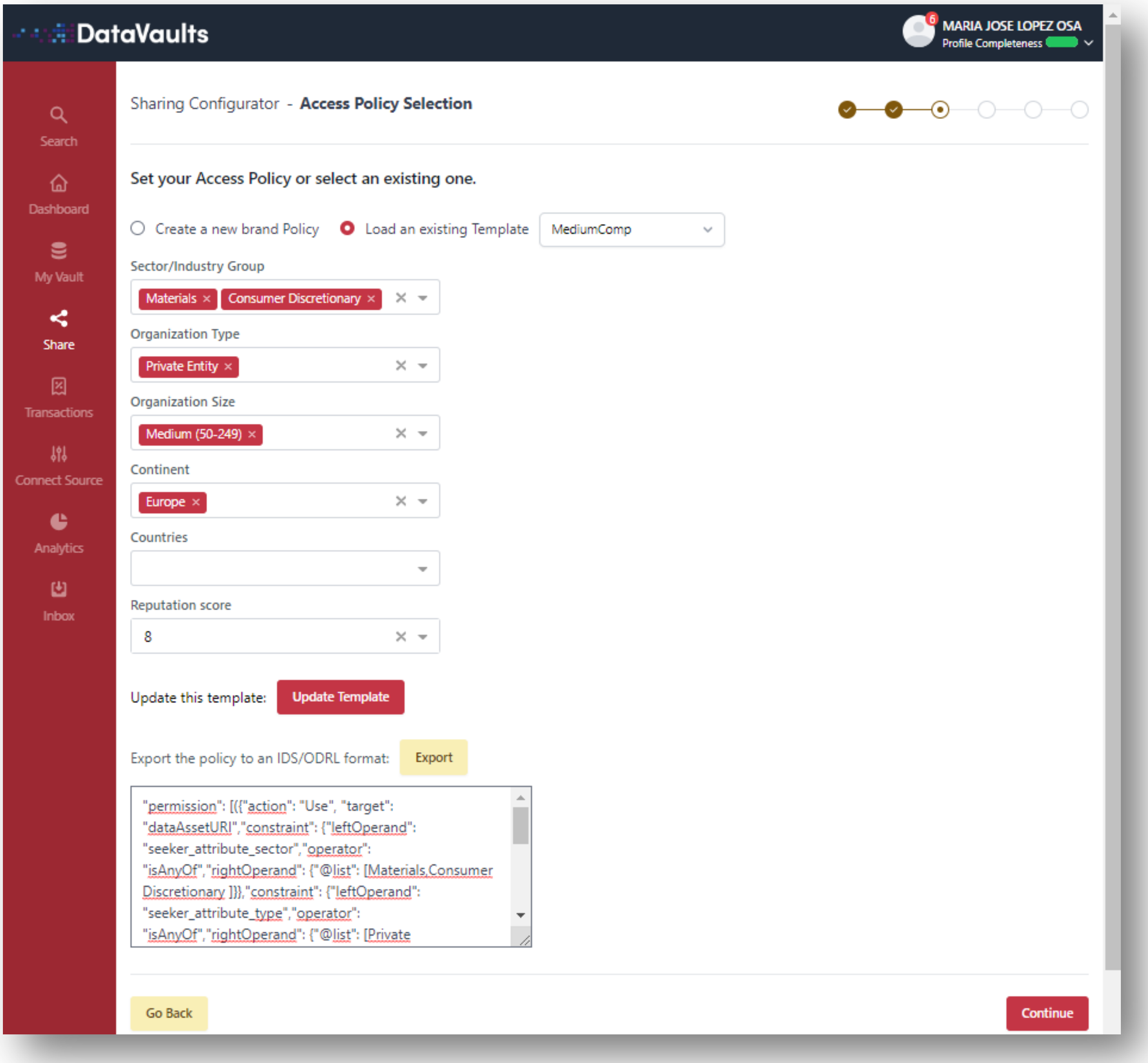
Countries: Belgium ×, Denmark ×, Germany ×

Reputation score: 9

Provide a name and description to save this policy as a template: smallCompanies | Access for small companie

Export the policy to an IDS/ODRL format:

Figure 18: The Access Policy Editor - New Policy



Sharing Configurator - **Access Policy Selection**

Set your Access Policy or select an existing one.

○ Create a new brand Policy ● Load an existing Template: MediumComp

Sector/Industry Group: Materials ×, Consumer Discretionary ×

Organization Type: Private Entity ×

Organization Size: Medium (50-249) ×

Continent: Europe ×

Countries:

Reputation score: 8

Update this template: Update Template

Export the policy to an IDS/ODRL format: Export

```
"permission": [({"action": "Use", "target":
"dataAssetURI","constraint": {"leftOperand":
"seeker_attribute_sector","operator":
"isAnyOf","rightOperand": {"@list": [Materials,Consumer
Discretionary ]}},"constraint": {"leftOperand":
"seeker_attribute_type","operator":
"isAnyOf","rightOperand": {"@list": [Private
```

Figure 19: The Access Policy Editor - Existing Policy

The main functionalities provided by this component in the current 0.5 version are:

- Create a new brand policy regarding the previously selected Data set.
- Load an existing policy template from a local repository. These policies are a set of reusable policies stored by the individual.
- Select the values for allowing access from a list of attributes of the seekers. The selection can be a list of allowed values.
- Save the conditions established in the current page locally as a reusable policy or update a specific one.
- Confirm the policy as the valid one and continue with the rest of the data sharing configuration.
- Transform (Export button) the policy created into an IDS/ODRL format. This functionality is a feature not to be used in the current DataVaults platform but to show the potential to adapt the policies to those policy formats.

2.8.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: As part of the Sharing Configurator component, it is built with the same VueJS3 framework and uses information coming from previous pages in this configurator.

2.8.2 Component Backlog

2.8.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the 0.5 version release, including the ones already implemented in previous releases. The latter have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID#	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_027	Access Policy Editor	DataVaults Personal App	Receive the identification of the Individual data set and load the access policies on the Access Policy Editor interface	The Individual can configure the policies for granting access to her data.
US_028	Access Policy Editor	Data Provider	load existing access policy templates for creating new policies	I can easily define the access policies that will apply to my data.
US_029	Access Policy Editor	Data Provider	create new templates from my policies	I can re-use it in the future.
US_030	Access Policy Editor	Data Provider	edit the access policies that apply to my data assets	I change the terms for providing access to my data
US_031	Access Policy Editor	Data Provider	finalise the policies configuration of a data sharing configuration.	these policies take effect once the sharing configuration is executed.
US_072	Access Policy Editor	DataVaults Personal App	Receive the identification of the Individual data set and load the access policies on the Access Policy Editor interface	The Individual can configure the policies for granting access to her data.
US_073	Access Policy Editor	Data Provider	edit the access policies that apply to my data assets	I change the terms for providing access to my data

US_076	Access Policy Editor	Data Provider	finalise the policies configuration of a data sharing configuration.	these policies take effect once the sharing configuration is executed.
US_073_1	Access Policy Editor	Data provider	Manage the access policies assigned to my data, allowing: - selection of a list of grantted values related to an attribute - reusable policies (templates) - load and modify the current policy	Manage the current policies allowing multiples values and reusable local policies for each user
US_073_2	Access Policy Editor	DataVaults personal App	Adapting policy representation to the DATAVAULTS model (DCAT + ODRL)	To be aligned with the data model defined in datavaults
US_074	Access Policy Editor	Data Provider	load existing access policy templates for creating new policies	I can easily define the access policies that will apply to my data.
US_075	Access Policy Editor	Data Provider	create new templates from my policies	I can re-use it in the future.

The features described above are already implemented and integrated in the available DataVaults Personal App. The development of the planned features could impact in them and they will be improved to be aligned with the evolution of the component.

The US_73_2 features has been implemented as a function for translating the policy into a more similar IDS/ODRL format. The Export button presents a textbox with this format.

2.8.2.2 Features planned for upcoming Releases

No more features are planned to be implemented for upcoming releases. The validation of the DataVaults Personal platform by the demonstrator could imply some adaptations or corrections in order to better cover their necessities. These modifications will be made within the project evolution, so possible additions may arise from the demonstrator validation in WP6.

3 WP3 Components Descriptions – Cloud Platform

The DataVaults Cloud Platform is a cloud service offering a single-entry point for Data Seekers. It includes the cloud-based infrastructure, and from the WP3 perspective, it includes the backend and frontend part of the following components:

- **Access Policy Engine** to control the access to specific data;
- **Persona Generator** to support data anonymization process;
- **Risk Management Monitor** to monitor and evaluate the risks related to the privacy exposure;
- **Data Stream & Contract Composer** to manage the lifecycle of the contracts;
- **Trusted DLT Engine and the Public and Private Ledgers** to facilitate the sealing of contracts on the side of the Individuals, as well as their compensation for assets that have been bought by Data Seekers;

The source code of the different components, which are open source, is provided in the following repository

https://www.gitlab.com/DataVaults

It is noted that for reasons of completeness, in case the scope and the technology background of a component have not been changed, the texts describing in the following sections are essentially the same as those presented in D3.2. In case there are changes, these are clearly marked in each description.

3.1 Access Policy Engine

Component's Concept Update from D3.2: None

Component Description: The Access Policy Engine is part of the DataVaults Cloud platform and responsible for analysing if a request of accessing data, made by the Data Seekers and managed by the Query Builder component in the DataVaults platform, will be granted.

The process followed by the Engine consists of comparing the current values of the attributes informed by Data Seekers and the values established as allowed by the data owners when configuring the Access Policies.

Once the Engine has performed the decision process, the response given from the component will consist of a Boolean parameter "granted", and in addition, if the access is not granted (false), a list of non-conformities will be sent to the caller. That outcome can be useful in case the Data Seeker wants to change the request or the attributes informed, if possible.

For providing the main functionality, that is to consider the request made by a Data Seeker allowed, the component needs to access to

- the information stored about the data seekers, including the one informed by them and the one calculated from previous experiences as a reputation score,
- the active policies of a specific data set in the DLT as part of the contracts.

This Engine does not provide a User interface to interact with it, considering it as an internal tool, part of the cloud platform and transparent to the users. This component exposes a function available for being called from the rest of the tools through an API. The implementation takes as input the IDs of the data sets requested and the ID of the Data Seeker.

3.1.1 Technology Background

Technology Background Update from D3.2: The Engine is integrated as a *Docker* element into the DataVaults Cloud Platform.

Technology Description: The Access Policy Engine is implemented in Java, using spring-boot framework and Swagger[6] for the specification and creation of the REST API.

[6] https://swagger.io/specification/

3.1.2 Component Backlog

3.1.2.1 Implemented Features (delivered in the 0.5 version Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_158	Access Policy Engine	DataVaults Cloud Platform	Identify the data involved in the request	APE accesses the policies associated to the selected data
US_159	Access Policy Engine	DataVaults Cloud Platform	Identify the attributes of a given Seeker	APE can use them for deciding on granting access or not.
US_160	Access Policy Engine	DataVaults Cloud Platform	Identify the information stored in the public ledger	APE retrieves any active sharing contracts, associated to the data asset
US_161	Access Policy Engine	DataVaults Cloud Platform	compare the Seeker's attributes with the access policies of the data	APE gives or denies access to the Seeker.
US_162	Access Policy Engine	Data Seeker	know why I was denied access, in case my request was denied	I can reconsider my profile attributes (ex. Submit documents to become a verified user)
US_164 (*)	Access Policy Engine	DataVaults Cloud Platform	gather the information about the access request resolution process after a Seeker's access request	the Contract Composer can register the information in a contract regarding the transaction.
US_163 (**)	Access Policy Engine	DataVaults Cloud Platform	call the ABE mechanism for sharing data, if the access is granted (by execution of the smart contract), and the data are encrypted under the ABE scheme	the second access privacy and security layer is activated, to provide access to the user.

US_165 (***)	Access Policy Engine	DataVaults Cloud Platform	trigger the component for creating the contract for an access request resolution process and provide the related information	the responsible component for the creation of the contract registers the access authorisation/deny process.

(*) Declared as obsolete and removed from the user stories list. At this moment the APEngine does not interact with the ledgers and does not handle more information than the policies, the seeker's attributes and the datasets IDs

(**) The query builder provides this functionality, once the APEngine returns the “granted” or “no granted” required.

(***) This is about registering an access request to the ledger, and this could be just a call with the APEngine results and the input to an API of the DLT, to record the output and the input of a APEngine process (all in 1 call, altogether). But in this current state of the component, it does not register the contract or the request. The APEngine just reads the policies and compares them with the attributes. The outcome of this process is a list of yes/no and the reasons why. At this moment, the query builder calls the APE and receives the outcome

3.1.2.2 Features planned for upcoming Releases

The features and so the functionalities provided by the Access Policy engine are stable since the previous version of the component, nevertheless, if the necessities of the project would evolve in this way, this feature would be implemented in future releases.

3.2 Risk Management Monitor and Dashboard

<u>Component's Concept Update from D3.2:</u> Since the previous reporting document (D3.2), the DataVaults Risk Management Monitor (including the part of the Personal App called Risk Management Dashboard related to the Data Owner user) has been further developed to allow privacy-related calculations based on GDPR fundamentals and the actual sharing aspects of each dataset.

With this latest release for V0.5, we allow the datasets that are ready to be shared to provide the corresponding information regarding anonymization, encryption and sharing rights to the Risk Management backend to calculate the privacy risk scoring. The scoring of each dataset is then displayed to the user of the Personal App (the Data Owner) while simultaneously providing an aggregation of the overall privacy risk to the cloud platform administrator.

The biggest update of the components is also the privacy risk calculation that is based on the idea of identifying and mapping Personal Identifiable Information (PII) that exists in the datasets and assessing if this sensitive information has been properly anonymized and encrypted, as well as how openly the dataset is shared.

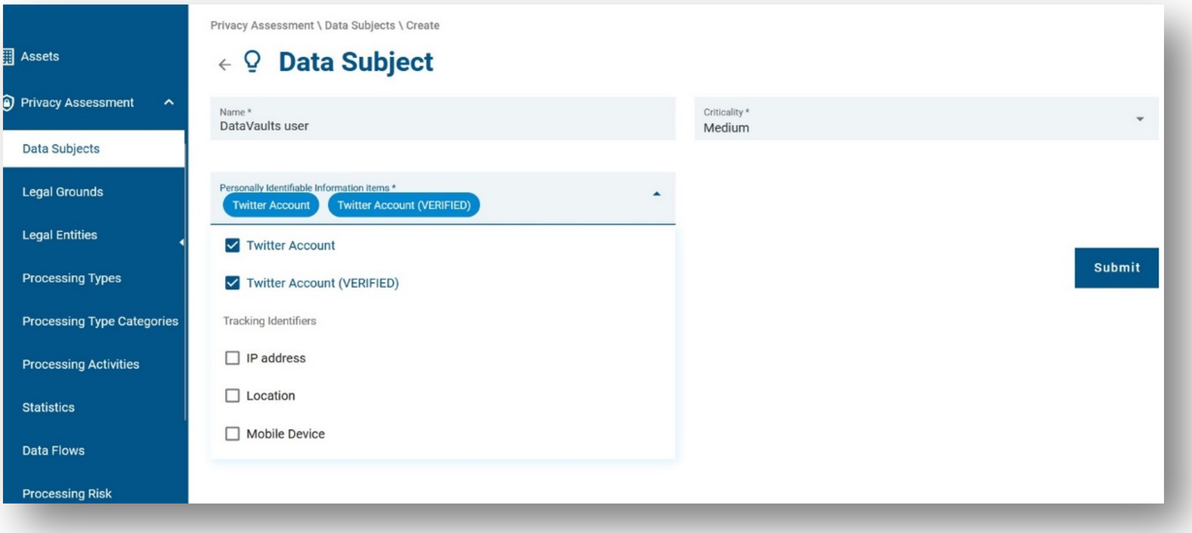
Component Description: The Risk Management Monitor is responsible for the evaluation of the risks related to privacy based on the identification of sensitive information that has been stored and shared through the datasets. For this to be achieved, the Risk Management Monitor has been integrated at the backend level with a) the Cloud Platform Backbone in order to have access to the users and their stored data, and b) the Personal App in order to retrieve the dataset sharing/anonymization/encryption configuration prior to the uploading of the dataset. Risk Management Monitor is part of the cloud platform and allows the administrator to make use of GDPR concepts to the datasets. At the same time, an API is provided to the Personal DataVaults App in order to visualize the privacy scores of the dataset for the end user.

More precisely, the Risk Management Monitor admin can use the provided web interface to define the DataVaults platform assets for the execution of risk assessment and to model the datasets that are shared through DataVaults to calculate the privacy risks. For the calculation of the overall risk assessment, an asset chain analysis is performed to estimate the impact that derives from the interconnections of the platform assets (and the uploaded datasets) and provide a single estimation for the overall impact.

In the following figures, we present the current version of the tool and explain how it is used to provide privacy assessment scoring for a dataset to be shared. The privacy assessment flow is based on the definitions and concepts of the GDPR that can be found in the article 4 of the GDPR legislation[7]. For this reason, the starting point is to select or create the data subjects of interest and include the PIIs as depicted in Figure 20.

[7] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02016R0679-20160504&from=EN#tocId7

As seen, PIIs can be multiple PIIs and can be selected through the available list of PIIs.



Privacy Assessment \ Data Subjects \ Create

Data Subject

Name * DataVaults user

Criticality * Medium

Personally Identifiable Information items * Twitter Account, Twitter Account (VERIFIED)

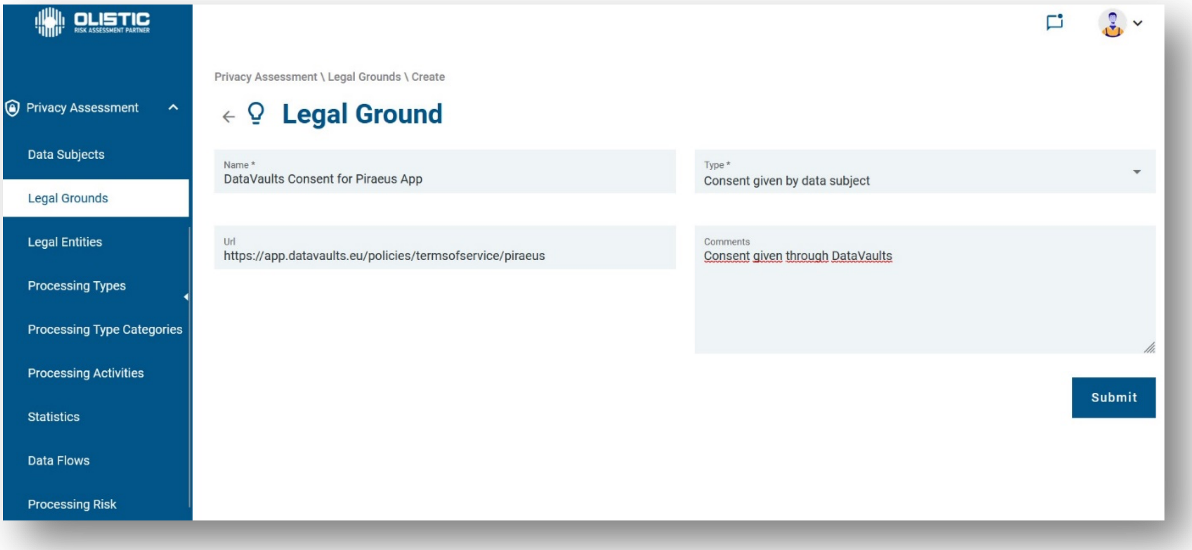
☑ Twitter Account
☑ Twitter Account (VERIFIED)

Tracking Identifiers

□ IP address
□ Location
□ Mobile Device

Figure 20: Definition of Data Subjects through the Risk Management Monitor

The data subjects are now part of the data processing (sharing) that can be performed in the data shared through DataVaults. For any data sharing or processing to be made, the appropriate consent shall be given; in the frame of DataVaults, this is performed when the user is providing consent when registering to the DataVaults through the Personal App side. Definition of the legal ground is provided as depicted in Figure 21.



Privacy Assessment \ Legal Grounds \ Create

Legal Ground

Name * DataVaults Consent for Piraeus App

Type * Consent given by data subject

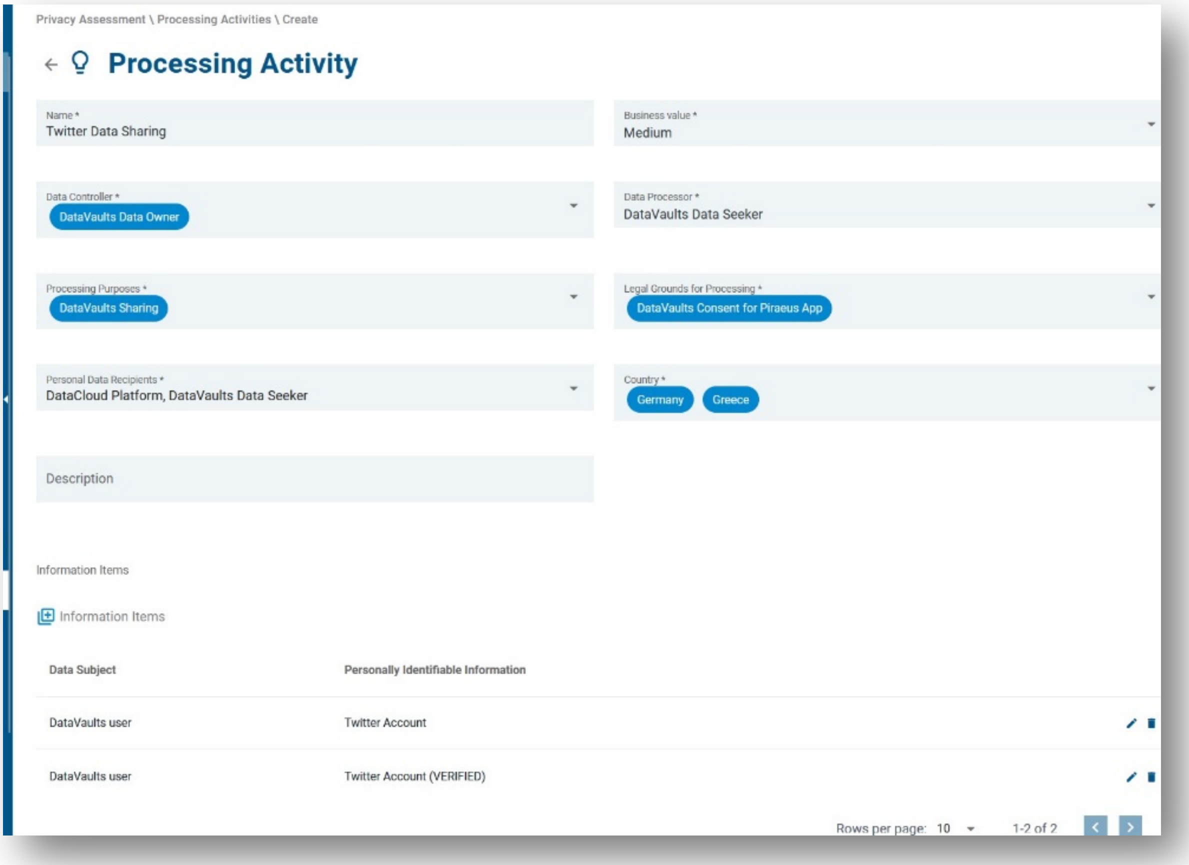
Url https://app.datavaults.eu/policies/termsofservice/piraeus

Comments Consent given through DataVaults

Figure 21: Definition of Legal Ground types from the Risk Management Monitor

Our tool also allows for properly defining the legal ground for processing the data concerning GDPR definitions, including the configuration of the Data Controller and Processor, the Data

Recipient and processing purpose. This information is used as seen in Figure 22, to model the processing activity.



Privacy Assessment \ Processing Activities \ Create

Processing Activity

Name * Twitter Data Sharing

Business value * Medium

Data Controller * DataVaults Data Owner

Data Processor * DataVaults Data Seeker

Processing Purposes * DataVaults Sharing

Legal Grounds for Processing * DataVaults Consent for Piraeus App

Personal Data Recipients * DataCloud Platform, DataVaults Data Seeker

Country * Germany Greece

Description

Information Items

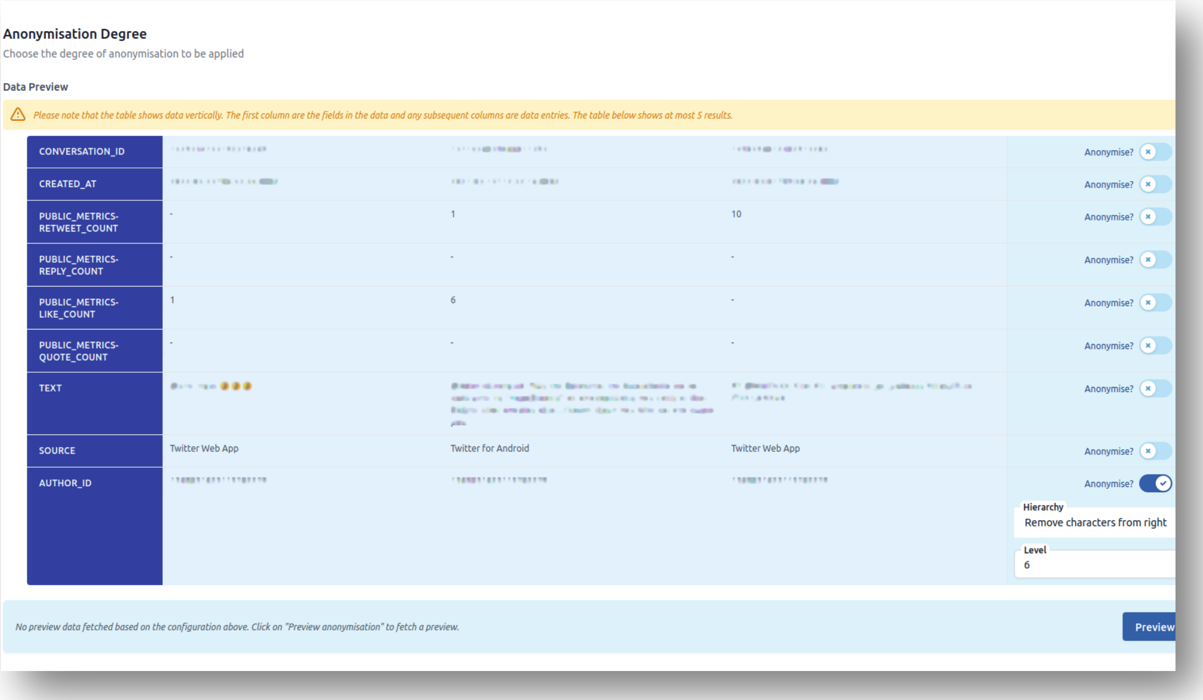
Information Items

Data Subject	Personally Identifiable Information
DataVaults user	Twitter Account
DataVaults user	Twitter Account (VERIFIED)

Rows per page: 10 1-2 of 2

Figure 22: Definition of processing activities to be assessed

The final and most important definition needed for the privacy assessment of the dataset is mapping the dataset columns to PIIs of the specific data subject. With these actions completed once for each dataset type, the sharing of the dataset by the user can be assessed. For this, we retrieve the information regarding the anonymization of each field, the encryption and sharing aspect through appropriate API calls.



Anonymisation Degree

Choose the degree of anonymisation to be applied

Data Preview

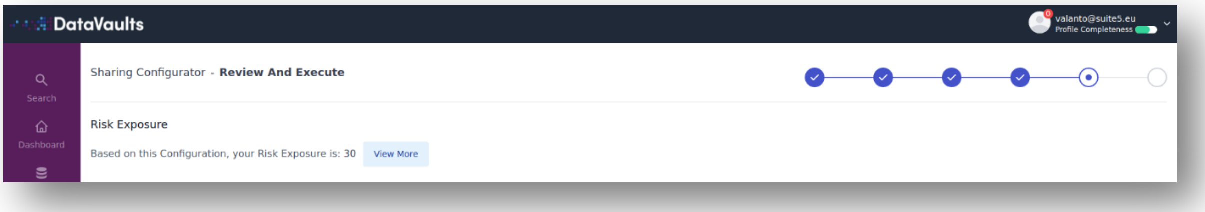
⚠ *Please note that the table shows data vertically. The first column are the fields in the data and any subsequent columns are data entries. The table below shows at most 5 results.*

CONVERSATION_ID	[redacted]	[redacted]	[redacted]	Anonymise? ☒
CREATED_AT	[redacted]	[redacted]	[redacted]	Anonymise? ☒
PUBLIC_METRICS-RETWEET_COUNT	-	1	10	Anonymise? ☒
PUBLIC_METRICS-REPLY_COUNT	-	-	-	Anonymise? ☒
PUBLIC_METRICS-LIKE_COUNT	1	6	-	Anonymise? ☒
PUBLIC_METRICS-QUOTE_COUNT	-	-	-	Anonymise? ☒
TEXT	[redacted]	[redacted]	[redacted]	Anonymise? ☒
SOURCE	Twitter Web App	Twitter for Android	Twitter Web App	Anonymise? ☒
AUTHOR_ID	[redacted]	[redacted]	[redacted]	Anonymise? ☑ Hierarchy: Remove characters from right Level: 6

No preview data fetched based on the configuration above. Click on "Preview anonymisation" to fetch a preview.

Figure 23: Anonymization information on the personal app

The result of the assessment process is the privacy risk exposure scoring of the dataset. Risk Management Dashboard is the view of Personal App that provides to the user the calculated privacy risk regarding the datasets that the user has added to the platform, and also warns the data owner of the risk exposure of a dataset that they are trying to share, as depicted on the top side of the figure below.



Sharing Configurator - **Review And Execute**

Risk Exposure

Based on this Configuration, your Risk Exposure is: 30

Figure 24: Privacy risk exposure on dataset sharing

3.2.1 Technology Background

Technology Background Update from D3.2:

For the v0.5, we have developed the backend and frontend for the functionalities presented above dedicated to the privacy assessment. Java and Vue.js has been used for this application, while the core model and backend of the the Privacy Assessment Tool (PAT)[8] of the *seCUre and pRivate hEalth data eXchange* (CUREX)[9] project, has been used as backend for the part of

[8] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/horizon-results-platform/26364

[9] https://curex-project.eu/

the risk assessment of the overall platform, allowing us to model the platform assets (e.g. servers), calculate to overall risk and provide to the administrator.

The Risk Management Monitor tool of is available at https://ra.datavaults.eu.

3.2.2 Component Backlog

3.2.2.1 Implemented Features (delivered in the v0.50 Release)

The features implemented at this point are supported at the backend level and the user interface.

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	Related Epic	User Story		
			As a <Role>	I want to <Action>,	so that <Reason>
US_166	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	assess the overall privacy exposure on the platform	I understand the current risk status
US_169	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	assess the privacy exposure of certain user, based on the updates of the data	risk metrics are provided to the administrator
US_077	Privacy Metrics Dashboard, Risk Management Monitor	Risk Assessment	Data Provider	know the risk related to the data I want share to the Cloud Platform	I fix any privacy-related issues in the data sharing configuration.

3.2.2.2 Features planned for upcoming Releases

The table below provides the list of features that remain in the backlog and are scheduled to be delivered in the final release by the integration work to be performed in WP5. With the main part of the privacy assessment and integration already completed, the focus will be to create new APIs and corresponding views that update the privacy risk and the overall risk and exposure values to the user.

ID #	Related Component	Related Epic	User Story		
			As a <Role>	I want to <Action>,	so that <Reason>
US_167	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	assess the privacy exposure of certain user,	risk metrics are provided to the administrator

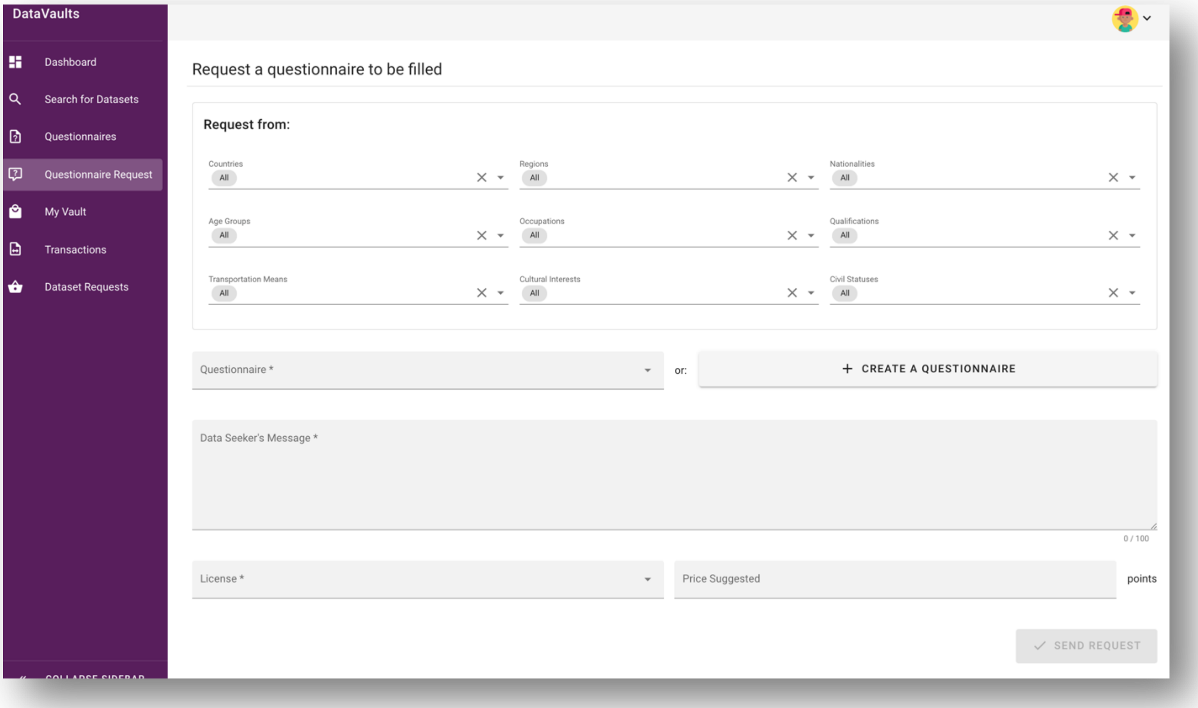
				based on the provided data	
US_168	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	update the privacy exposure of certain user, based on the downloads of data by a Data Seeker	risk metrics are provided to the administrator
US_170	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	assess the privacy exposure of certain user, based on the provided data	risk metrics are provided to the user
US_171	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	update the privacy exposure of certain user, based on the downloads of data by a Data Seeker	risk metrics are provided to the user
US_172	Risk Management Monitor	Risk Assessment	DataVaults Cloud Platform	assess the privacy exposure of certain user, based on the updates of the data	risk metrics are provided to the user
US_078	Privacy Metrics Dashboard, Risk Management Monitor	Risk Assessment	Data Provider	view a graphical representation of the risk values related to my shared data	I modify the sharing configuration or make the data completely unavailable for sharing.
US_079	Privacy Metrics Dashboard, Risk Management Monitor	Risk Assessment	Data Provider	know the privacy exposure of specific datasets that I have shared in the Cloud	I modify the sharing configuration or make the data completely unavailable for sharing.
US_080	Privacy Metrics Dashboard, Risk Management Monitor	Risk Assessment	Data Provider	the privacy metrics of my data to be updated based on the downloads performed by a Data Seekers	I modify the sharing configuration or make the data completely unavailable for sharing.
US_081	Privacy Metrics Dashboard, Risk Management Monitor	Risk Assessment	Data Provider	know the overall privacy metrics of my user account	I modify the sharing configuration or make the data completely unavailable for sharing.

3.3 DATASTREAM AND CONTRACT COMPOSER

Component's Concept Update from D3.2: None

Component Description: This component is responsible for the execution of data trading contracts whenever a Data Seeker is willing to acquire a dataset. The component is therefore used to request the instantiation of a data trading contract when an asset has an already fixed price, thus executing the contract as present in the ledger and recording a relevant transaction. This is an operation not visible to the user, as the contract is executed immediately without allowing any modification.

Additionally, this component handles data sharing requests when a Data Seeker selects to acquire either an already shared but not with a fixed price dataset, or new information from a Data Owner through a questionnaire builder, as shown in the figures below.



Request a questionnaire to be filled

Request from:

Countries: All	Regions: All	Nationalities: All
Age Groups: All	Occupations: All	Qualifications: All
Transportation Means: All	Cultural Interests: All	Civil Statuses: All

Questionnaire *

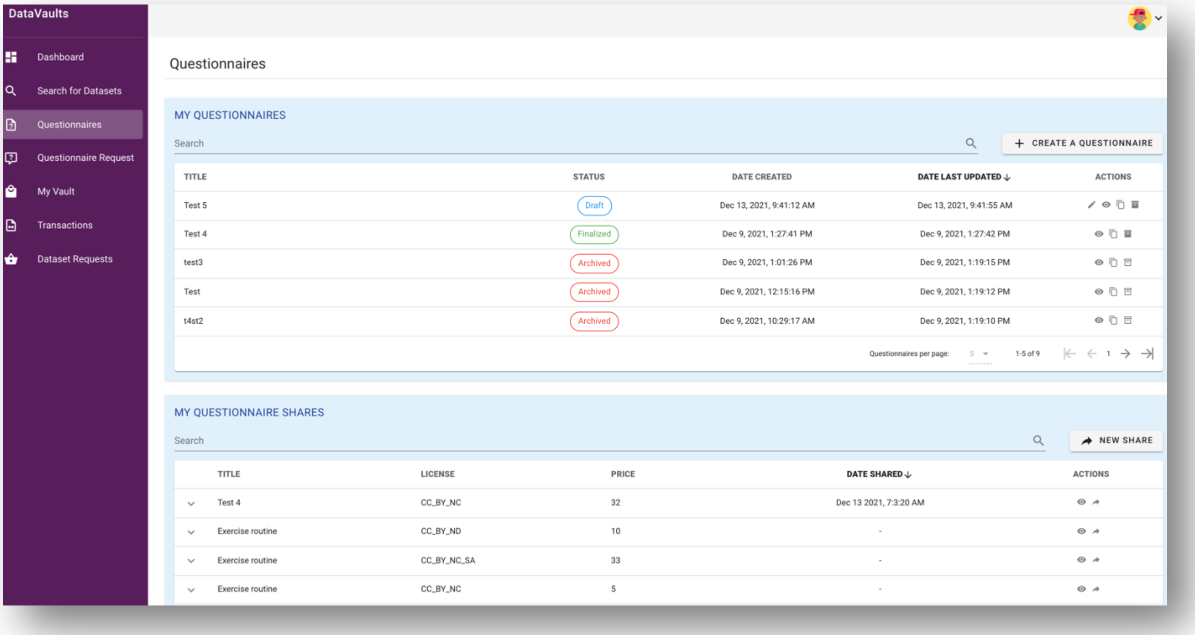
Data Seeker's Message *

0 / 100

License *

Price Suggested points

Figure 25: Screenshot of Building a Questionnaire Technology Background



Questionnaires

MY QUESTIONNAIRES

TITLE	STATUS	DATE CREATED	DATE LAST UPDATED ↓	ACTIONS
Test 5	Draft	Dec 13, 2021, 9:41:12 AM	Dec 13, 2021, 9:41:55 AM	
Test 4	Finalized	Dec 9, 2021, 1:27:41 PM	Dec 9, 2021, 1:27:42 PM	
test3	Archived	Dec 9, 2021, 1:01:26 PM	Dec 9, 2021, 1:19:15 PM	
Test	Archived	Dec 9, 2021, 12:15:16 PM	Dec 9, 2021, 1:19:12 PM	
t4st2	Archived	Dec 9, 2021, 10:29:17 AM	Dec 9, 2021, 1:19:10 PM	

Questionnaires per page: 5 1-5 of 9 1

MY QUESTIONNAIRE SHARES

TITLE	LICENSE	PRICE	DATE SHARED ↓	ACTIONS
Test 4	CC_BY_NC	32	Dec 13 2021, 7:3:20 AM	
Exercise routine	CC_BY_ND	10	-	
Exercise routine	CC_BY_NC_SA	33	-	
Exercise routine	CC_BY_NC	5	-	

Figure 26: Screenshot of designed and shared Questionnaires

3.3.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The technology for implementing this component is VueJS2 and storage of the information that is necessary to build the questionnaire is done in the Postgres database which is located in the cloud-based infrastructure.

The communication between this component and the Data Request Service resolver is done via RabbitMQ.

3.3.2 Component Backlog

3.3.2.1 Implemented Features (delivered in the v0.50 release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>	so that <Reason>
US_186	DataStream & Contract Composer	Data Seeker	have a contract created every time I purchase data through DataVaults	the transaction and terms are recorded for logging and auditing purposes.
US_187	DataStream & Contract Composer	Data Seeker	have the contract stored in a secure manner	I am sure that it will not be tampered with.
US_188	DataStream & Contract Composer	Data Provider	have a contract created every time my data are purchased through DataVaults	the transaction and terms are recorded for logging and auditing purposes.
US_189	DataStream & Contract Composer	Data Provider	have the contract stored in a secure manner	I am sure that it will not be tampered with.
US_190	DataStream & Contract Composer	Data Seeker	be able to compose a draft contract for a request for data that are not yet available through DataVaults	I can make an offer to an Individual for data assets.
US_191	DataStream & Contract Composer / Notification System	Data Provider	receive a data sharing request with a predefined sharing configuration	I can quickly accept or decline the request
US_192	DataStream & Contract Composer	Data Provider	automatically share the requested asset in case I have accepted the request	I skip the sharing configuration step.
US_193	Notification System	Data Seeker	I want to receive a notification based on the outcome of a request	I can find out whether I possess the data or not

3.3.2.2 Features planned for upcoming Releases

All identified features have been delivered. If new user stories are identified during development and demonstrator validation (WP6), they will be included in subsequent deliverables of WP5.

3.4 Policy-compliant Blockchain Infrastructure and DLT Engine

Component's Concept Update from D3.2: None. The focus in this updated release of the overall Blockchain infrastructure is on providing the pending functionalities w.r.t dynamic user addition in the policy-compliant Blockchain ecosystem (once the user has been successfully registered to the overall Blockchain platform through the implemented Identity Manager) so as to enable their secure interactions and transactions with the distributed ledgers. Furthermore, the necessary APIs and interfaces for integrating a user's Personal Wallet to the created ledgers was performed so as to be able to securely transfer the necessary credits, during a data trading operation, from the Data Seeker to the DataVaults platform and then to the Data Owner.

Component Description: The DLT Engine is responsible for the recording and sharing of all operational-related data so as to facilitate any data trading actions initiated by (authorized) Data Seekers. It provides all necessary functionalities, capturing both the data sharing transactions (data uploaded by the Data Owners) as well as data trading transactions, through the construction and execution of the necessary smart contracts enabling authenticated and authorized entities/users to interoperate in a trusted and secure manner. Certificate-based authentication is managed by the DataVaults Identity Manager (leveraging the KeyCloak technology) who, in turn, distributes attribute-based user tokens to the internal Context Broker (as part of the DLT Engine) for enabling attribute-based access control to the on-chain stored data blobs. Access control policies are created through the Data Sharing Configurator and the Access Policies Editor and then managed through smart contract functions that are invoked prior to providing a user access to requested data. While the DLT Engine has been configured to work in tandem with Keycloak, acting as an OAuth2 provider that distributes data access tokens to users and validates these tokens, DataVaults also supports the integration of electronic Identification, Authentication and Trust Services (eIDAS) that offer Verifiable Credentials (VCs) issued by EU-certified service providers in the form of JWT tokens. This enables DataVaults to adopt the Single-Sovereign Identity (SSI) paradigm providing a trusted identity management mechanism capable of asserting both the claimed identify a user (either Data Owner and/or Data Seeker) but also the validity of the issued attributes. For the latter, DataVaults has also enhanced the structure of JWT tokens (as currently issued by existing eIDAS-enabled issuers) to consider any additional attributes that might be required by a data sharing policy.

Once the user has been authenticated successfully, the DLT Engine (through the internal GoQuorum and Tessera Nodes acting as Blockchain Peers – as described in D3.2, DataVaults is based on the use of the GoQuorum Blockchain infrastructure) provides the necessary set of APIs enabling the secure interactions with the (permissioned) private and public ledgers as it pertains to the: (i) recording of a data blob - uploaded by a Data User – accompanied with the respective data sharing policy, (ii) the querying of an access control and sharing policy for a specific piece of data to be provided to a Data Seeker prior to allowing a further data trading, and (iii) recording and auditing of the required consent prior to the exchange/sharing of data to a requesting Data Seeker. In this context, the use of the ledgers is to ensure the data and event traceability across the entire data market and to be able to provide the required data security, user privacy and ledger security properties.

3.4.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The current implementation of the DataVaults policy-compliant Blockchain infrastructure is based on the use of the open-source Quorum technology. As was described in D3.2, we have extended the novel concept of "**Multi-Tenancy via Multiple Private States**" so as to enhance the scalability of the overall framework by enabling multiple users using the same GoQuorum and Tessera nodes (as Blockchain Peers) without any breach on their privacy.

All smart contracts and the respective CreateSharingPolicy, AddDataBlob, QueryPolicy functions have been written in Go language. This also enables DataVaults to support the secure execution of such functionalities through Peers equipped with a trusted component (such as a TPM) allowing the establishment of decentralized systems where each node/peer can first validate the integrity of the data source prior to recording the data on the ledger; i.e., DataVaults is working on an updated version of the Trusted Software Stack (TSS) written in Go for enabling the direct execution of a smart contract function through the trusted component.

Finally, as aforementioned, in its current version, the DLT Engine has been equipped to interact with the DataVaults Keycloak identity and authentication manager supporting the integration of both OAuth-based certificates as well as the use of attribute-based verifiable credentials issued by an eIDAS-enabled VC Issuer.

3.4.2 Component Backlog

3.4.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_202	Private Ledger	Data provider	a record to be created each time my data are shared to the Cloud	the activity is securely logged.
US_204	DataVaults Private Brokerage Engine	Data Provider	have the agreed amount of currency transferred to my wallet	I receive the agreed compensation for my data.

US_203	Private Ledger	Data Provider	have transaction privacy	my personal information (configuration, account value) is hidden from other users.
US_207	Open Ledger	DataVaults Cloud Platform	restrict the access to the ledger	the functionalities are only provided to authorized Data Seekers.
US_208	Open Ledger	Data Seeker	have an open ledger account created when I register	I can make currency transactions to purchase data
US_209	Open Ledger	Data Provider	sell my asset only to verified data seekers	only trusted individuals receive my data.
US_210	Open Ledger	Data Seeker	transact the agreed amount of currency from my wallet to DataVaults	I can purchase a data asset from a Provider.
US_211	Open Ledger	Data Seeker	display the value of my wallet account	I can evaluate the worth of the previous sharing activities

3.4.2.2 Features planned for upcoming Releases

All identified features of the initial platform requirements have been developed w.r.t the management and recording of all data sharing and trading operations.

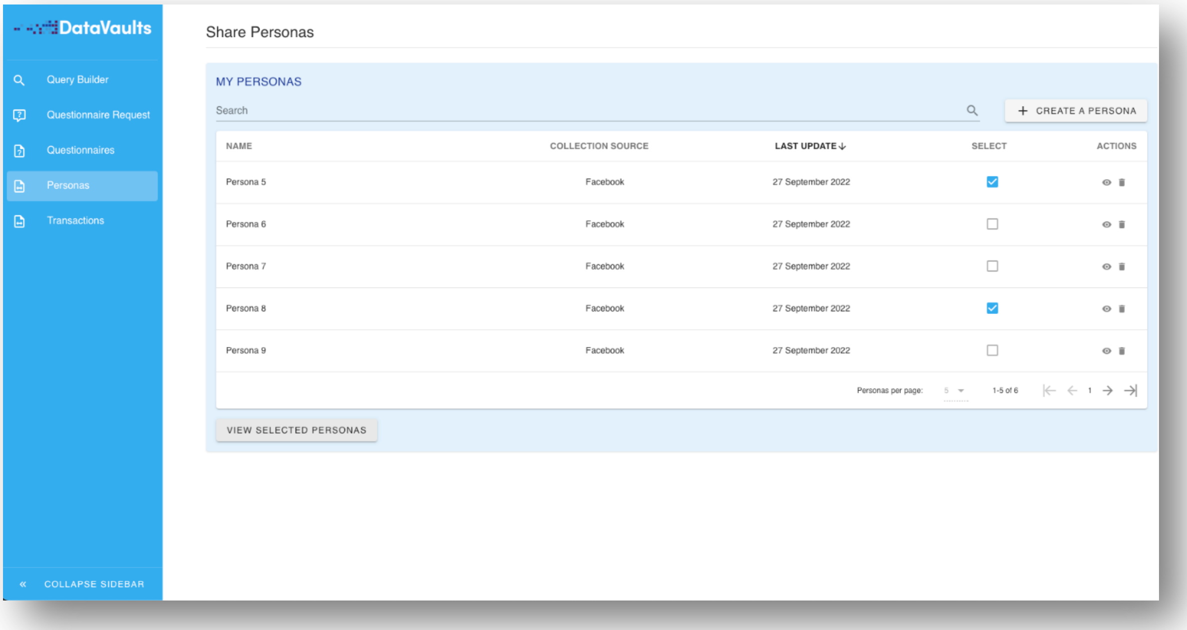
In its final release, the consortium will explore the enhanced of the DLT Engine with additional APIs for the better interaction with the Personal Wallet component of a user so as to be able to support the (privacy-preserving) display and evaluation of all data trading activities been performed by all registered users.

3.5 Persona Generator

<u>Component's Concept Update from D3.2:</u> None

Component Description: The DataVault's vision of a persona abandons the idea of traditional market research in favour of a modern data-driven approach. A **marketing persona** draws a picture of who your target audience is. Marketing personas are based on market **research via focus groups and interviews**, typically to represent the largest group you plan to target. Our data-driven approach is based on aggregation of data and the use of analytics (statistical and machine learning) to draw out insights and generalised characteristics that are representative of the data providers, from whom the aggregated data has been sourced. This process consists of taking a dataset of users and separating them into groups with similar demographic details. Insights from these groups can be found and used to generate a user persona. In this case, each persona would be a representative of a user group.

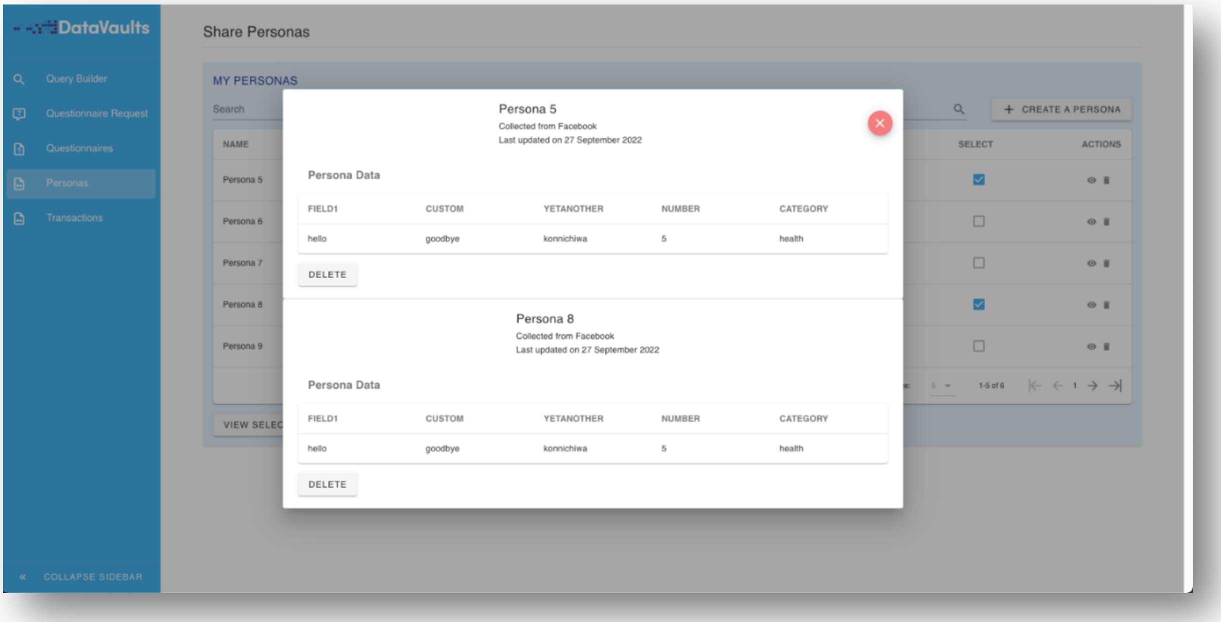
This process of persona generation is particularly of note to DataVaults as it would allow users to opt in to sharing their data as part of a group as opposed to a simple 1-to-1 anonymization process. By using the dataset to generate the insights incorporated into the persona, each user's identity would be concealed as no direct link can be made back to the dataset using the persona.



NAME	COLLECTION SOURCE	LAST UPDATE ↓	SELECT	ACTIONS
Persona 5	Facebook	27 September 2022	☑	
Persona 6	Facebook	27 September 2022	□	
Persona 7	Facebook	27 September 2022	□	
Persona 8	Facebook	27 September 2022	☑	
Persona 9	Facebook	27 September 2022	□	

Personas per page: 5 1-5 of 6 1

Figure 27: Persona Generator - Main View



Persona 5
Collected from Facebook
Last updated on 27 September 2022

Persona Data

FIELD1	CUSTOM	YETANOTHER	NUMBER	CATEGORY
hello	goodbye	konnichiwa	5	health

DELETE

Persona 8
Collected from Facebook
Last updated on 27 September 2022

Persona Data

FIELD1	CUSTOM	YETANOTHER	NUMBER	CATEGORY
hello	goodbye	konnichiwa	5	health

DELETE

Figure 28: Persona Generator - Detail View

3.5.1 Technology Background

Technology Background Update from D3.2: None

Technology Description: The persona generator currently consists of a single python script containing multiple methods which carry out the primary function of the persona generator. We are using a combination of Dask[10] and Pandas[11] to fulfil our data processing needs such as encoding non-numerical values in the dataset, preparing the dataset for evaluation by the machine learning algorithm and extracting insights from the clustered data. Sckit-learn is then used to apply the mean-shift algorithm to the dataset which will separate the dataset into clusters by enhancing the dataset with a new column labelling each data-point based on the cluster it belongs to. The enhanced dataset is then passed through a series of methods that iterates through each cluster extracting statistical insights from those clusters. These are then stored in a dictionary.

The workflow for the persona generator is as documented in the UML diagram in Figure 29.

[10] https://dask.org/

[11] https://pandas.pydata.org/docs/getting_started/index.html

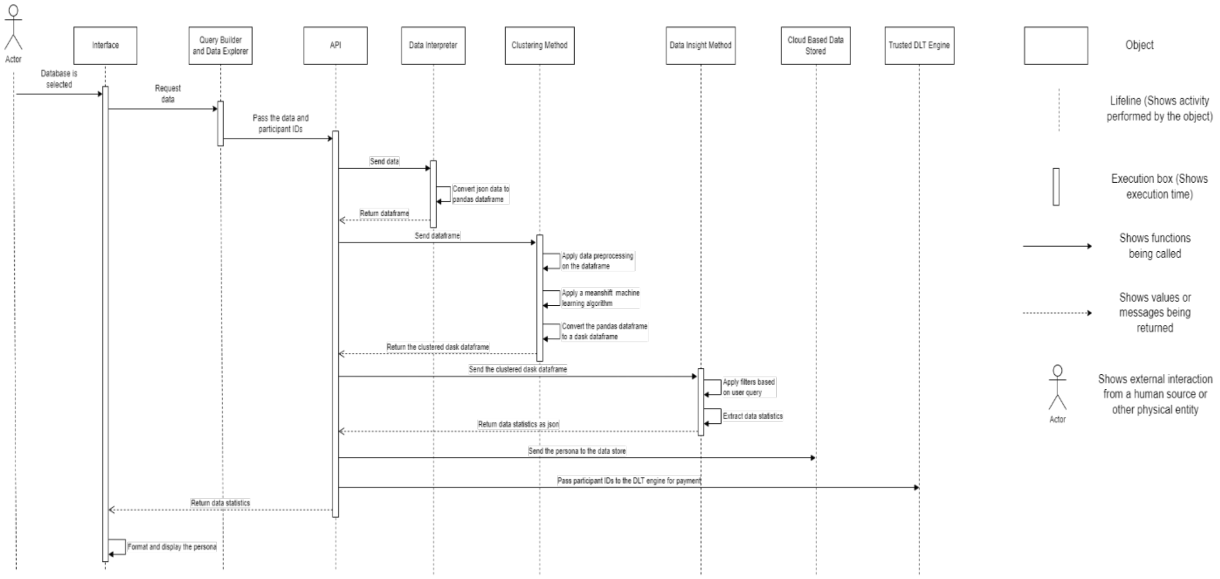

Actor
Interface
Query Builder and Data Explorer
API
Data Interpreter
Clustering Method
Data Insight Method
Cloud Based Data Stored
Trusted DLT Engine
Database is selected
Request data
Pass the data and participant IDs
Send data
Convert json data to pandas dataframe
Return dataframe
Send dataframe
Apply data preprocessing on the dataframe
Apply a meanshift machine learning algorithm
Convert the pandas dataframe to a dask dataframe
Return the clustered dask dataframe
Send the clustered dask dataframe
Apply filters based on user query
Extract data statistics
Return data statistics as json
Send the persona to the data store
Pass participant IDs to the DLT engine for payment
Return data statistics
Format and display the persona
Object
Lifeline (Shows activity performed by the object)
Execution box (Shows execution time)
Shows functions being called
Shows values or messages being returned
Actor
Shows external interaction from a human source or other physical entity

Figure 29: Workflow of Persona Generator

3.5.2 Component Backlog

3.5.2.1 Implemented Features (delivered in the v0.50 Release)

The table below provides the list of the features that have been delivered as part of the v0.50 release, as well as the list of features that have been part of the Alpha and the Beta release. All features (new and old ones) have been subject to optimisations and bug fixing, following the continuous integration and agile development approach followed by the project.

ID#	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_236	Persona Creation	Data Provider	share my data provided it is used for persona generation	the level of privacy is maintained
US_138	Insight_dask.py	Data Provider	be able to share insights as part of a group	my underlying personal data is not shared
US_237	Persona Creation	Data Provider	have my data assets that reside on the DataVaults Cloud Platform to be utilised in variants of persona creation	I remain completely anonymous
US_238	Persona Creation	Data Seeker	commission adjustments to a persona to a Data Analyst	he can adjust it accordingly to exactly match my specific needs

US_239	Persona Creation	Data Seeker	commission a persona to a DataVaults Analyst	he can create a persona that exactly match my specific needs
US_240	Persona Creation	DataVaults Analyst	execute the Persona generation queries on the data stores on the platform	I can form the personas
US_149	Persona Generator	Data Seeker	see how a Persona has evolved over time	I identify any existing trends.
US_241	Persona Searching	Data Seeker	search for personas and have access to a wide range of personas generated	I can be more economical than using the raw data.
US_242	Persona Searching	Data Seeker	see a list of all personas available in the DataVaults Cloud Platform	I can generate more specific understandings
US_148	Persona Generator	Data Seeker	be notified about the progress of my request for a Persona	I keep track of its current status.
US_142	Persona Generator	DataVaults Data Scientist	mix personas	I create integrated Personas with enriched content.
US_154	Persona Generator	DataVaults Data Scientist	Delete a Persona	It is no longer available

3.5.2.2 Features planned for upcoming Releases

No further features are planned – additional updates will be based primarily on resolving any issues that arise from ongoing testing and validation and improvements made based on specific user feedback

4 CONCLUSIONS AND NEXT STEPS

This document describes the components from the DataVaults architecture assigned to WP3. The components provide the core functionalities to enable secure data sharing and access, as well as privacy and trust preservation.

Based on the release roadmap and development plan of the project, the implementation of most components has been finished. The remaining user stories are completed as part of the verification and integration activities of WP5. Resulting from the demonstrator validation performed in WP6, improvements and minor features might be added to the already defined components. Such alterations are also conducted and documented as part of WP5.

This document is the last iteration of the series and marks the end of work package 3.