



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

D4.2

Data Sharing, Value Generation and Intelligence Bundles - Version 2

Editor(s)	Sotiris Koussouris
Lead Beneficiary	Suite5 Data Intelligence Solutions
Status	Draft
Version	1.0
Due Date	31/12/2021
Delivery Date	29/12/2021
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	WP4 - Multitude Trusted Intelligence Bundles for Personal Data Insights Generation
Deliverable	D4.2 - Data Sharing, Value Generation and Intelligence Bundles - Version 2
Editor(s)	Suite5 – Sotiris Koussouris
Contributor(s)	FRAUNHOFER – Torben Jastrow, Simon Dutkowski ATOS – Tomás Pariente Lobo, Iván Martínez Rodríguez, Javier Villazán UBITECH – Giannis Ledakis, Konstantinos Oikonomou
Reviewer(s)	TECNALIA – María José López MIWENERGIA – Borja Molina

Abstract	This deliverable is provided as an auxiliary document for the delivery of the code of the components of the WP4 activities, providing a high-level description of those
Disclaimer	<p>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.</p> <p>© Copyright in this document remains vested with the DataVaults Partners</p>

Executive Summary

D4.2 is the second version of deliverable D4.1, whose objective is to present the status of the implemented service bundles of the project coming out of WP4.

The present deliverable is provided as an auxiliary document for the delivery of the code of the different components of DataVaults (which is the actual content of 4.2), as these are grouped under a deliverable of type OTHER.

The components that are presented in this deliverable are the following:

- The Data Fetcher and Transformation
- The Query Builder
- The Data Explorer
- The SSE Engine
- The Data Analytics Playground
- The Edge Analytics Engine

Table of Contents

1	Introduction	6
1.1	Document structure	6
1.2	Relation To Other WPs/Tasks	6
2	WP4 Components Descriptions	7
2.1	Data Fetcher and Transformation	8
2.1.1	Technology Background	8
2.1.2	Component Backlog	8
2.2	Query Builder.....	11
2.2.1	Technology Background	12
2.2.2	Component Backlog	12
2.3	Data Explorer	15
2.3.1	Technology Background	15
2.3.2	Component Backlog	15
2.4	SSE Engine.....	17
2.4.1	Technology Background	17
2.4.2	Component Backlog	17
2.5	Data Analytics Playground.....	19
2.5.1	Technology Background	20
2.5.2	Component Backlog	21
2.6	Edge Analytics.....	23
2.6.1	Technology Background	23
2.6.2	Component Backlog	23
3	Conclusions and Next Steps	25
	References.....	26

List of Figures

Figure 1: Query Builder - Search View Screenshot	11
Figure 2: Query Builder – Advanced Search View Screenshot.....	12
Figure 3: Data Explorer Screenshot.....	15
Figure 4: SEAS User Interface	19
Figure 5: SEAS Deployment User Interface	19
Figure 6: Mlflow	20
Figure 7: Apache Superset.....	20

1 INTRODUCTION

The present deliverable is provided as an auxiliary document for the delivery of the code of the different components of DataVaults that are the direct outcome of WP4 (for M24 of the project), as these are grouped under a deliverable of type OTHER.

It is noted that for reasons of completeness, that in case the scope and the technology background of a component has not been changes, the texts describing those in this section and in the sections following are essentially the same as those presented in D4.1 [1], In case there are changes, these are marked under a subsection in each description.

1.1 DOCUMENT STRUCTURE

This document starts with an introduction in Section 1, and Section 2 provides an overview for the different components that have been the focus of the work until M24 of the project. For each one, a short description is provided, including mock-up screens or real-life screenshots (depending on whether the component has a UI and based on the progress made in the UI design), and the technology stack for each implementation is briefly provided. Moreover, for each component, the updated backlog of user stories coming out of WP5 is provided, identifying the user stories that have been completed so far in terms of development and indicating the ones which remain at the backlog and that will be considered for implementation in the upcoming releases.

Finally, section 3 of the documents provides a short concluding statement.

1.2 RELATION TO OTHER WPs/TASKS

The D4.2 deliverable, as the second version of the WP4 series of deliverables, is closely related to the WP5 and WP3 deliverables. The different components and the user stories that drive their implementation are defined in WP5, and as such input is received from WP5.

In addition, the outcomes of D4.2 are provided as input to WP5, on two different levels:

- Firstly, the code of WP4 is provided as ready to integrate components into WP5 which will in turn handle the integration of those into the overall DataVaults solution alongside with their testing.
- Secondly, as the development of the different components and the tackling of the different user stories is following an agile approach in WP4, feedback is provided to WP5 which impacts the overall design and evolution of the architecture and might also impact the user stories.

Finally, as WP4 provides components that highly interact with those of WP3, a bidirectional communication channel between those WPs is established, and D4.2 is considering the outputs of D3.2 [2], while it also provides feedback to be used as input for the components that are developed under WP3 (and in this release are submitted as deliverable D3.2)

2 WP4 COMPONENTS DESCRIPTIONS

This section provides a high-level overview of the different components that are developed under WP4 of the project, according to the DoA and the latest developments of the DataVaults architecture, which mostly concern data management and analytics that are to be provided by the DataVaults Infrastructure.

In more detail, the components that are presented in the next subsections are the following:

- The Data Fetcher and Transformation, which is tasked with the collection of data from individuals, and as such is a component housed in the DataVaults Personal App
- The Query Builder is a component operated by the Data Seekers that allows querying the Data Store of the cloud-based platform. It presents Data Seekers with a data catalogue and the contents of the data store (which they can access)
- The Data Explorer allows a Data Seeker to browse their own Data Spaces, where the assets they have already bought reside. This is a component which is not depicted in the first version of the architecture, but is going to be part of the second version
- The SSE Engine, which is tasked with performing the necessary activities for encrypting and decrypting the various data based on Symmetric Searchable Encryption scheme
- The Data Analytics Playground that is used by Data Seeker for designing and running analytics jobs based on the datasets they have acquired
- The Edge Analytics Engine, which is a lightweight analytics implementation, offered to Data Owners to get insights on the data that reside in their Personal App

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

<https://www.gitlab.com/DataVaults>

2.1 DATA FETCHER AND TRANSFORMATION

Component's Concept Update from D4.1: None

Components Description: The Data Fetcher & Transformer component utilizes a modular micro-service architecture. The fetcher module can periodically check an API which is selected by a user (out of a list of available APIs in the Personal App which can connect to external sources) to find data and collect them. At this stage, only a small subset of predefined APIs is supported. The data then is transformed into the DataVaults metadata data model and exported into a MongoDB¹ for further usage.

In its current form, it is possible to configure the schedule for a preconfigured data source and to start and stop the collection schedule. The Data Fetcher & Transformer does not have a user interface, as all configuration is done via the DataVaults Personal App interface. The configurations for the Data Fetcher and Transformer regarding the user management will reside on the backend on a special multiuser scheme.

2.1.1 Technology Background

Technology Background Update from D4.1: None

Technology Description: The Data Fetcher and Transformer is programmed in Java and Kotlin code. It uses Eclipse Vert.X² as web framework. The resulting micro-services are based on *piveau consus*³.

2.1.2 Component Backlog

This section provides the full set of features that belong to the backlog, splitting those in the ones already implemented, and the ones that are pending implementation.

2.1.2.1 Implemented Features (delivered in the Alpha and Beta Release)

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. New user stories that were not part of the initial backlog provided in D4.1 are marked with an asterisk (*) next to their ID.

¹ <https://www.mongodb.com/>

² <https://vertx.io/>

³ <https://www.piveau.de/en/>

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
Features provided as part of the Beta Release				
US_33	Data Fetcher & Transformation	Data Provider	manually select the specific data that will be collected from the connected data source	I have control on which data assets are collected by my DataVaults Personal app and exclude any data that I don't want to be transferred from the data source.
Features Implemented by the Alpha Release (and optimised if needed in Beta Release)				
US_34	Data Fetcher & Transformation	Data Provider	execute a data collection configuration	I can collect data from the connected data source.
US_36	Data Fetcher & Transformation	Data Provider	add a collection schedule to a collection configuration	I can configure a recurring collection process for the specific data source (e.g., once a day, once a week etc.)
US_38	Data Fetcher & Transformation	Data Provider	pause a collection schedule	I can stop my DataVaults personal App from collecting data from a specific data source for a time period and if I change my mind later, restart data collection using the same configuration
US_40	Data Fetcher & Transformation	Data Provider	remove a collection schedule from a saved configuration	the collection no longer happens automatically and recurrently.
US_45	Data Fetcher & Transformation	DataVaults Personal App	I want to transform the fetched data into the DataVaults data model	I can ensure compliance with the available DataVaults schemas as well as the necessary data quality levels to achieve maximum usefulness and usability.

2.1.2.2 Features planned for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4. It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which 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_35	Data Fetcher & Transformation	Data Provider	save a data collection configuration	I can reuse it in the future.
US_37	Data Fetcher & Transformation	Data Provider	configure how data from the intervals between recurring collections will be handled	I can configure the "density" of the data that are collected in my DataVaults personal App.
US_39	Data Fetcher & Transformation	Data Provider	edit a saved data collection configuration	I can change the selected data, collection schedule etc.
US_41	Data Fetcher & Transformation	Data Provider	delete a data collection configuration	I completely discard a collection configuration I no longer need
US_42	Data Fetcher & Transformation	DataVaults Personal App	link different datasets from the collection tasks of one account	I know that these refer to the same Individual
US_43	Data Fetcher & Transformation	DataVaults Personal App	semantically enrich the collected data	I can retrieve similar data from different sources with different representations.
US_44	Data Fetcher & Transformation	DataVaults Personal App	do a Quality Check over the fetched data	I know, how to transform the data into the DataVaults data model.

2.2 QUERY BUILDER

Component's Concept Update from D4.1: None

Component's Description: The Query Builder is the main facility used by Data Seekers for searching within the cloud-based platform to retrieve datasets, which are shared by the different data owners. In the backend, the Query Builder uses a Triple Store for Linked Data to store, retrieve and search information about the data that is retrieved via the frontend, while the core of the dataset is stored in a non-relational database.

This approach enables a hybrid solution that allows semantic search and interoperability of data at the metadata level, providing functionalities such as those of a data catalogue. At the same time the core datasets can be indexed (if required) and served in a faster and more efficient manner to the requested component, directly from the non-relational data store.

For this purpose, apart from the usual APIs to communicate with the non-relational data store, an additional micro-service will provide a RESTful API that translates frontend requests into SPARQL queries.

The frontend interface of the Query builder is based on VueJS to match it with the UX of the rest of the DataVaults cloud-based platform, and various operations that have to do with calling APIs from other components are handled by the frontend. In this respect, data retrieval is performed by the backend APIs and in turn communicating their results with the APIs of the Access Policy Engine, as access to the queried data shall be resolved by the Access Policy Engine, by utilizing as input information from the ledger. Moreover, this component will interact also with the SSE engine, to allow searching over encrypted data in the later releases of the platform.

Below follow some screenshots of the component, as to be delivered in the beta release.

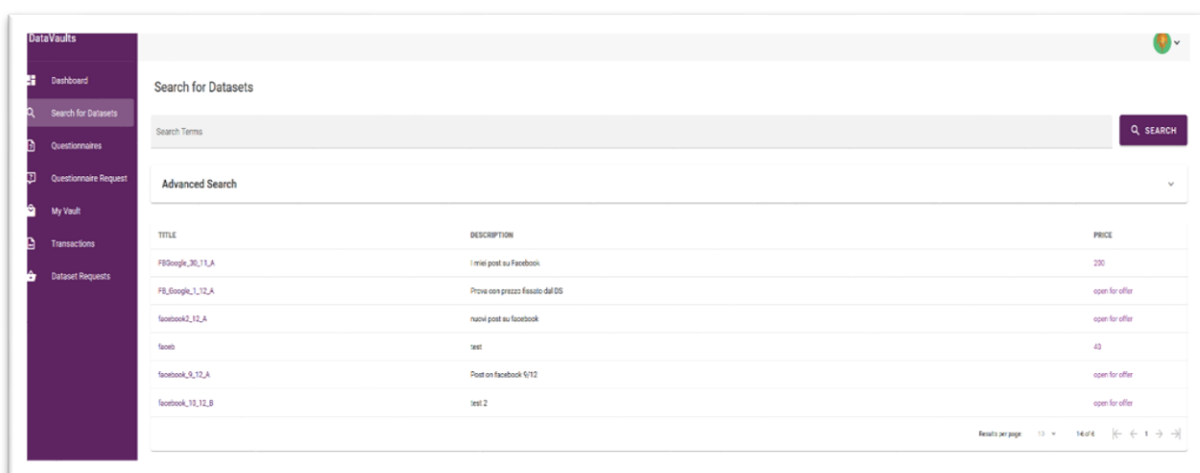


Figure 1: Query Builder - Search View Screenshot

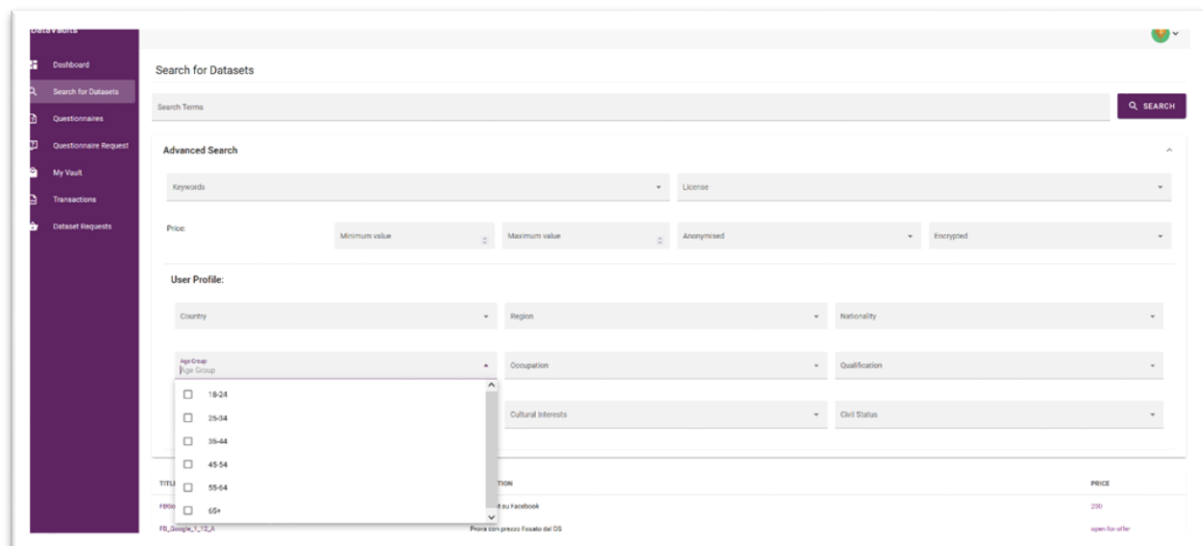


Figure 2: Query Builder – Advanced Search View Screenshot

2.2.1 Technology Background

Technology Background Update from D4.1: None

Technology Description: The backend of the Query Builder is written in Java and Kotlin and uses the Eclipse Vert.X web framework. For data storage, an OpenLink Virtuoso⁴ server is used. The Backend is based on *piveau hub*. The frontend of the Query builder is provided using the VueJS2⁵ framework.

2.2.2 Component Backlog

2.2.2.1 Implemented Features (delivered in the Alpha and Beta Release)

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. New user stories that were not part of the initial backlog provided in D4.1 are marked with an asterisk (*) next to their ID.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
Features provided as part of the Beta Release				
US_214	Query Builder	Data Seeker	filter the enlisted data assets by choosing from the available values in various categories	I can limit the list of enlisted data assets to those that match my

⁴ <http://virtuoso.openlinksw.com>

⁵ <https://vuejs.org/>

			(such as domain - e.g., health -, format -e.g., json-, anonymisation level - e.g., none, digital twin, persona-, language - e.g., GR, DE,- etc.)	needs.
Features Implemented by the Alpha Release (and optimised if needed in Beta Release)				
US_213	Query Builder	Data Seeker	purchase/acquire one of the enlisted data assets	I get access to the data.
US_215	Query Builder	Data Seeker	search the data assets available in the DataVaults Cloud based on their name, using keywords or the exact title	I can easily find data assets that I know their names or their name indicates they could be useful for me.
US_216	Query Builder	Data Seeker	search the data assets available in the DataVaults Cloud based on metadata (e.g., description, author, domain, tags) using keywords	I can easily find data assets that their metadata match my needs.
US_217	Query Builder	Data Seeker	search the data assets available in the DataVaults Cloud based on metadata such as age group, price, etc. using ranges	I can easily explore data assets that their metadata match my needs, even when I do not want to provide an exact value, but I am interested in a group.
US_220	Query Builder	Data Seeker	see a list of all the data assets matching to my search criteria	I can explore the results of my search.
US_221	Query Builder	Data Seeker	select one of the enlisted data assets and see all the information about this data asset that is available for preview	I can see if this data asset contains the information I need.

2.2.2.2 Features planned for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4.

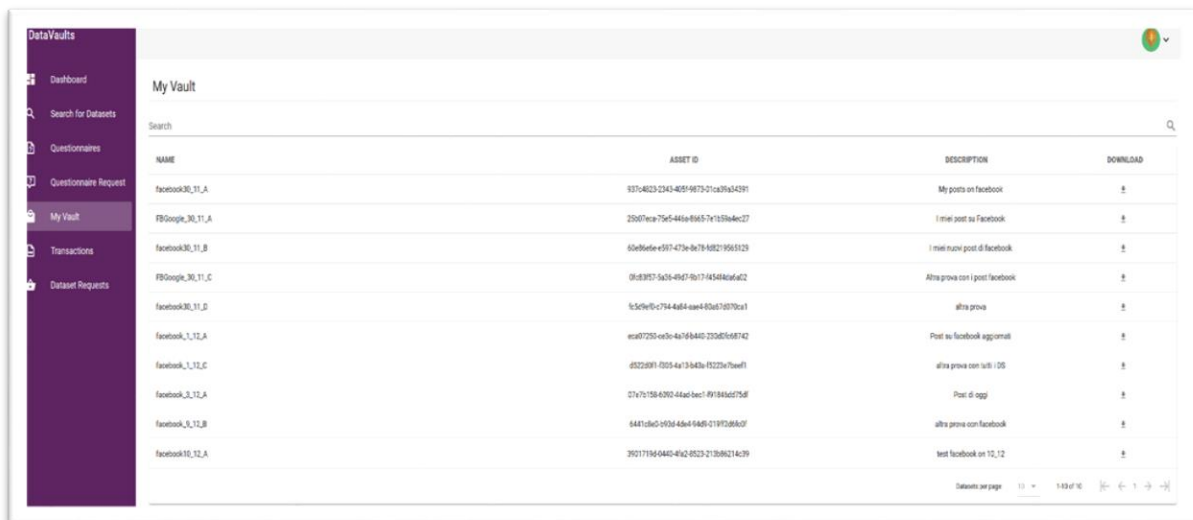
It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which 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_218	Query Builder	Data Seeker	search the data assets available in the DataVaults Cloud based on their content, using keywords	I can easily find data assets that contain information that could be useful for me.
US_219	Query Builder	Data Seeker	search the data assets available in the DataVaults Cloud Platform using complex queries, i.e., by defining more than one criterion in my search	I can find data assets that match to various aspects.
US_222	Query Builder	Data Seeker	save queries	I can use them again in the future.
US_223	Query Builder	Data Seeker	load and edit saved queries	I can add or remove any search criteria that I do not need for my new search.
US_224	Query Builder	Data Seeker	delete saved queries	I can remove a query I no longer need

2.3 DATA EXPLORER

Component's Concept Update from D4.1: None

Component's Description: Initially part of the larger component called “Query Builder and Data Explorer”, this component is the one which allows Data Seekers to browse into their own Data Space and retrieve, delete or forward for analysis the data assets they have already bought.



The screenshot shows the 'DataVaults' interface with a sidebar menu on the left containing 'Dashboard', 'Search for Datasets', 'Questionnaires', 'Questionnaire Request', 'My Vault', 'Transactions', and 'Dataset Requests'. The 'My Vault' section is active, displaying a table of data assets. The table has columns for NAME, ASSET ID, DESCRIPTION, and DOWNLOAD. The data is as follows:

NAME	ASSET ID	DESCRIPTION	DOWNLOAD
facebook30_11_A	937c4023-2343-405f-9879-01ca9a5439f1	My posts on facebook	⬇
FBGoogle_30_11_A	25a07eca75e5446a89a57a165ba4ec07	I met post su Facebook	⬇
facebook30_11_B	60e9f4ee4587473e9a7848219565329	I met nuovi post di facebook	⬇
FBGoogle_30_11_C	09cd39575a3649e75b1714548a6af4c2	Altra prova con i post facebook	⬇
facebook30_11_D	1c3c9a10c7144a84aee480a67d070a1	altra prova	⬇
facebook_1_12_A	eca07293-ea3e-4a716940-220d51a6742	Post su facebook aggiornati	⬇
facebook_1_12_C	492289f1-0054-4a713443a-f022e75ae01	altra prova con tutti i DS	⬇
facebook_3_12_A	07a7d158430241a4a1ee7f8f38a6d754f	Post di oggi	⬇
facebook_3_12_B	6441da6c103a43a4194b0119f0388a0f	altra prova con facebook	⬇
facebook10_12_A	3921716644404fa24023213086214c39	test facebook on 12_12	⬇

At the bottom right of the table, it says 'Datasets per page' with a dropdown menu set to '10 of 10' and navigation arrows.

Figure 3: Data Explorer Screenshot

2.3.1 Technology Background

Technology Background Update from D4.1: The main technology background for this component has been shifted to Java from Python, to allow for better interoperability with other components of the backend which are also Java applications

Technology Description: The Data Explorer is written in Java, using VueJS2 in the frontend.

2.3.2 Component Backlog

2.3.2.1 Implemented Features (delivered in the Alpha and Beta Release)

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. New user stories that were not part of the initial backlog provided in D4.1 are marked with an asterisk (*) next to their ID.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
Features provided as part of the Beta Release				
US_232	Data Explorer	Data Seeker	upload my own data as files	I can use them for analytics.
US_234	Data Explorer	Data Seeker	search for specific data assets using keywords	I can easily explore my data assets.
Features Implemented by the Alpha Release (and optimised if needed in Beta Release)				
US_230	Data Explorer	Data Seeker	browse through the assets I have already purchased/stored in my storage container	I can use it in the analytics containers
US_231	Data Explorer	Data Seeker	select from available options (e.g., 'Group by source', 'Group by domain', etc.) to view my data	I can group my data.

2.3.2.2 Features planed for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4. It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which 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_233	Data Explorer	Data Seeker	upload my own data through an API	I can use them for analytics.
US_235	Data Explorer	Data Seeker	filter data assets based on metadata values	I can easily explore my data assets.

2.4 SSE ENGINE

Component's Concept Update from D4.1: None

Component's Description: The SSE Engine of DataVaults aims to encrypt metadata of files and then allow users to search on top of those encrypted metadata, adopting the approaches used in the ASCLEPIOS project, where metadata and data content from selected files are encrypted using a symmetric key and the cloud-based service can perform queries over the store ciphertext, without knowing the keys to decrypt them. SSE at the first stage will be applied only at the metadata level, as this is considered more efficient from a resource point of view, and keys of SSE will be protected through the ABE Engine, so that only Data Seekers with the appropriate attributes will be in a position to search over the encrypted metadata. As such, integration with the ABE engine is planned, and will be delivered in the next stages of the project.

During this period, work has focused on analysing the possible implementation options and building the data flow and the appropriate infrastructure upon which the SSE Engine will be developed, and development of the key features has started. A first version of this component is expected to be released in the Beta release of the platform.

The SSE engine does not have a visual interface and is fully a back-end service used by the Query Builder, operating without the Data Owner or Seeker needing to configure anything, as keys are automatically issued and used, once a Data Owner chooses to provide his/her data in an encrypted manner to DataVaults.

2.4.1 Technology Background

Technology Background Update from D4.1: None

Technology Description: The technology used in the SSE engine is based on Python code, reusing libraries developed in the frame of the ASCLEPIOS H2020 project

2.4.2 Component Backlog

2.4.2.1 Implemented Features (delivered in the Alpha and Beta Release)

No features have been completed to be released in the Beta Version.

2.4.2.2 Features planned for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4. It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which 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_03	Encryption/Decryption Engine	Data Provider	be sure that the encryption keys used to secure my data will not be made available to anyone else when I decide to share my data through the Cloud	my data that are protected with cryptographic techniques are not at risk.
US_107	DV Cloud Platform Backbone. Encryption / Decryption	Data Provider	have selected data assets that reside on the DataVaults Cloud Platform to be stored encrypted	only I and Data Seekers with the appropriate rights can get access to them.
US_108	DV Cloud Platform Backbone. Encryption / Decryption	Data Provider	be sure that the decryption keys provided to Data Seekers for getting access to my data assets, are configured so that they enforce the sharing terms in place (e.g., expire after a period, are only for one use etc.)	the sharing of my data assets complies to the sharing terms I have defined.
US_109	DV Cloud Platform Backbone. Encryption / Decryption	Data Seeker	be able to search over the encrypted data assets available over the DataVaults Cloud	I can find any data assets of interest.

2.5 DATA ANALYTICS PLAYGROUND

Component's Concept Update from D4.1: None

Component's Description: The Secure Analytics Playground (SEAS) is a module inside DataVaults platform that offers the possibility of running different algorithms over the shared data in a secure way.

The SEAS is a module developed for being used by the Data Seekers however Data Providers could utilize it to gain insights into their own data. The module is divided into two subcomponents. The first subcomponent "Service Analytics Host" is in charge of defining the features of the setup and configuration of the Playground and it also has a UI to manage and define these characteristics. The user will be able to deploy the playground on a defined server or local, for example on their own laptop. The Data Seeker will have the Playground and Visualization modules in an isolated environment to create their own models or algorithms or to select the ones provided by the platform

The second subcomponent "Playground & Visualization Host" will allow the Data seeker to execute and track the Machine Learning models. The tracking capability allows the user to decide which one of these will be stored in the Cloud, sent to the Visualization module, or both.



Figure 4: SEAS User Interface

Figure 5: SEAS Deployment User Interface

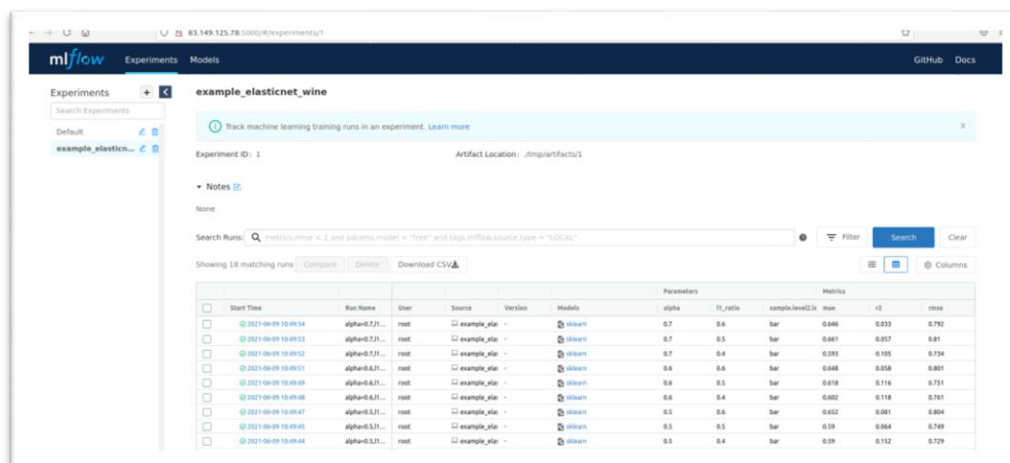


Figure 6: MLflow

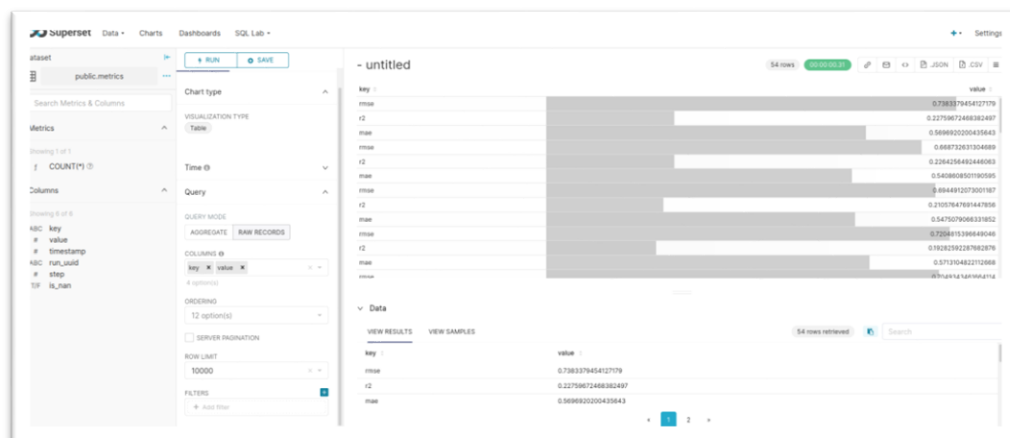


Figure 7: Apache Superset

2.5.1 Technology Background

Technology Background Update from D4.1: None

Technology Description: The "Service Analytics Host" will use technologies such as Play for the User Interface, Java for the backend setup and Docker for the deployment. Ansible will allow us to execute the different tasks associated with the deployment on various platforms (e.g. local or on a defined server).

The "Playground & Visualization Host" will be based on MLflow⁶ to execute and track the Machine Learning models, PostgreSQL⁷ to storage the results of the models and Apache Superset to visualize the results

⁶ <https://mlflow.org/>

⁷ <https://www.postgresql.org/>

2.5.2 Component Backlog

2.5.2.1 Implemented Features (delivered in the Alpha and Beta Release)

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. New user stories that were not part of the initial backlog provided in D4.1 are marked with an asterisk (*) next to their ID.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
Features provided as part of the Beta Release				
US_176	Secure Analytics Playground	Data Seeker	run an analytics task over my requested data	I can get my needed information from DataVaults
US_179	Secure Analytics Playground	Data Seeker	be able to run my own algorithms on the cloud	I can easily use specialized algorithms over the gathered data
US_181	Secure Analytics Playground	Data Provider	see graphics and visualizations of the algorithms results	I can gain some insights about my data
US_184	Secure Analytics Playground	Data Seeker	save a visualisation in my storage container	I can use it in the future
Features Implemented by the Alpha Release (and optimised if needed in Beta Release)				
US_173	Secure Analytics Playground	Data Seeker	be able to configure an analytics task	I can perform analytics and gain insights from my data.
US_174	Secure Analytics Playground	Data Seeker	select the data on which I want to run an analysis	I can generate intelligence out of the data purchased
US_175	Secure Analytics Playground	Data Seeker	configure and execute a visualisation task	I can gain visual insights from my data.
US_177	Secure Analytics Playground	Data Seeker	run aggregation/filtering algorithms on the cloud	the amount of data I need to download is reduced
US_178	Secure Analytics Playground	Data Seeker	see graphics and visualizations of the analytics task	I avoid running visualisations in-house

2.5.2.2 Features planed for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4. It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which are market with an asterisk (*) next to their ID.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
US_180	Secure Analytics Playground	Data Provider	run algorithms over my own data and compare it to other "open" data in the platform	I can gain some insights about my data
US_182	Secure Analytics Playground	Data Seeker	download the results of an analysis to my storage container	I can reuse them in the future
US_183	Secure Analytics Playground	Data Seeker	download the results of an analysis to my local device	I can use them offline
US_185	Secure Analytics Playground	Data Seeker	download a visualisation to my local device	I can use it offline
US_250*	Secure Analytics Playground	Data Seeker	fetch into the SEAS engine data assets that reside in my Vault (of the cloud platform)	I can perform analytics on those data

2.6 EDGE ANALYTICS

Component's Concept Update from D4.1: None

Component's Description: The Edge Analytics component will be offered as a component of the DataVaults Personal App tasked to provide Data Owners with ready-made visualisations concerning their data as well as some analytics regarding their activity over DataVaults.

These will be done by identifying specific data sources for which predefined analytics will be set up, so that the users can view relevant metrics without asking them to interact with the component. Apart from that, the component will also include an analytics library and engine, allowing more experienced users to define a specific analytics workflow and execute that flow on their side to generate intelligence out of their data.

2.6.1 Technology Background

Technology Background Update from D4.1: Due to the nature of analytics to be executed, a shift to the technology stack of Pandas and scikit-learn has been decided, as these are lightweight and can easier fit to the concept of the analytics to run at the side of the data owners.

Technology Description: The Edge Analytics component shall be based on the Pandas⁸ library and scikit-learn⁹. The frontend of the component is implemented using the VUEJS3 framework

2.6.2 Component Backlog

2.6.2.1 Implemented Features (delivered in the Alpha and Beta Release)

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. New user stories that were not part of the initial backlog provided in D4.1 are marked with an asterisk (*) next to their ID.

ID #	Related Component	User Story		
		As a <Role>	I want to <Action>,	so that <Reason>
Features provided as part of the Beta Release				
US_82	Edge Analytics Engine	Data Provider	select an analytics algorithm from a list of available algorithms to perform analysis on my data	I can get specific insights regarding my data and their progress.

⁸ <https://pandas.pydata.org/>

⁹ <https://scikit-learn.org/>

US_83	Edge Analytics Engine	Data Provider	select a specific set of data to perform analysis	I can get insights regarding data from specific sources, during a specific period.
US_86	Edge Analytics Engine	Data Provider	select a specific set of data to visualise	I can get better visual insights regarding data from specific sources, during a specific period.
US_88	Edge Analytics Engine	Data Provider	generate data assets from the results of the analysis of my data	I can handle them as any other personal data asset: store them, export them, and share them through DataVaults.

2.6.2.2 Features planned for upcoming Releases

The table below the features that remain in the backlog and are scheduled to be delivered in the following releases of WP4. It is noted that as the project is based on an agile design and development approach, certain user stories might have been removed from the initial backlog provided in D4.1 as they are either covered by new features or are deemed as obsolete as the project's concept evolves, while new one might have surfaced, which 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_84	Edge Analytics Engine	Data Provider	view on my device the results of the analytics task over my data	I can easily understand my activity and progress in relation to factors such as time.
US_85	Edge Analytics Engine	Data Provider	select a visualisation method from a list of available methods such as graphs, timelines, pie charts	I can easily understand my data and their progress through their visual presentation.
US_87	Edge Analytics Engine	Data Provider	view on my device the visualisation results of the defined visualisation task	I can easily understand my activity and progress in relation to factors such as time.
US_89	Edge Analytics Engine	Data Provider	generate data assets from the results of the visualisation of my data	I can handle them as any other personal data asset: store them, export them and share them through DataVaults.

3 CONCLUSIONS AND NEXT STEPS

This document provides a high-level overview of the different components coming out of WP4 and that are described by the DataVaults architecture as components that are working mostly in the data management and analytics tasks of the overall system.

At the time of submitting this document, those components have already progressed with regards to their implementation, following the development plan of the project, which is derived from the MVP and of the needs of the project which stem out of the requirements that should be covered by the Beta release of the project.

As such, the current implemented features are provided to WP5 for the integration tasks, while work in WP4 will continue towards further developing the different components. In this direction, it is expected that new features for each component might surface, as the overall architecture will evolve, while there is also the case that certain parts of the architecture might also call for the introduction of new components in this WP.

An updated version of this document is due to M32 of the project, as deliverable D4.3.

REFERENCES

- [1] DataVaults Consortium, "D4.1 Data Sharing Value Generation and Intelligence Bundles - Version 1," 2021.
- [2] DataVaults Consortium, "D3.2 Security, Privacy and Trust Bundles - Version 2," 2021.