Project acronym: EVIDENT

Project title:

bEhaVioral Insights anD Effective eNergy policy acTions

Deliverable 6.4

Datahub Services of EVIDENT platform

Document Control Page

Deliverable Name	Datahub Services of EVIDENT platform
Deliverable Number	D6.4
Work Package	WP6
Associated Task	Task 6.2 System Development
Covered Period	M18-M30
Due Date	31st of May 2023
Completion Date	31st of May 2023
Submission Date	2nd of June 2023
Deliverable Lead Partner	SID
Deliverable Author(s)	Konstantinos Kyranou (SID), Theocharis Saoulidis (SID), Zisis Batzos (SID), Eleftherios Fountoukidis (SID), Ioannis Hadjigeorgiou (SID), Thomai Karamitsou (SID), Anastasios Lytos (SID), Athanasios Tziouvaras (Bi2S), Christos Doumanopoulos (CERTH), Karypidis Paris (DUTH), Panagiotis Sarigiannidis (UOWM), Anna Triantafyllou (UOWM), Athanasios Liatifis (UOWM), Georgios Fragulis (UOWM),
Version	1.0

Dissemination Level		
PU	Public	X
CO	Confidential to a group specified by the consortium (including the Commission Services)	

Document History

Version	Date	Change History	Author(s)	Organisation
0.1	January 19, 2023	ToC Formulation	Konstantinos Kyranou, Theocharis Saoulidis, Zisis Batzos, Eleftherios Fountoukidis, Ioannis Hadjigeorgiou, Thomai Karamitsou, Anastasios Lytos	SID
0.2	Feb 24, 2023	Section 3	Konstantinos Kyranou, Panagiotis Sarigiannidis, Anna Triantafyllou, Athanasios Liatifis, Georgios Fragulis, Theocharis Saoulidis,	SID, UOWM

			Zisis Batzos, Eleftherios Fountoukidis, Ioannis Hadjigeorgiou, Thomai Karamitsou, Anastasios Lytos	
0.3	March 10, 2023	Section 2	Athanasios Tziouvaras	Bi2S
0.4	April 01, 2023	Initial version of section 5	Karypidis Paris, Christos Ntoumanopoulos	DUTH, CERTH
0.5	April 15, 2023	Final version of section 5	Karypidis Paris, Christos Ntoumanopoulos	DUTH, CERTH
0.6	April 28, 2023	Section 4	Konstantinos Kyranou, Karypidis Paris, Theocharis Saoulidis, Zisis Batzos, Eleftherios Fountoukidis, Ioannis Hadjigeorgiou, Thomai Karamitsou, Anastasios Lytos	SID, DUTH
0.7	May 15, 2023	Finalisation, updated content added	Konstantinos Kyranou, Theocharis Saoulidis, Zisis Batzos, Eleftherios Fountoukidis, Ioannis Hadjigeorgiou, Thomai Karamitsou, Anastasios Lytos	SID
0.8	May 29, 2023	Internal Review	Grigorios Koutantos, Athanasios Tziouvaras	PPC, Bi2S
0.9	May 29, 2023	Quality and Risk Review	Dimosthenis Ioannidis	CERTH
1.0	May 31, 2023	Final version submitted to the European Commission	Konstantinos Kyranou, Theocharis Saoulidis, Zisis Batzos, Eleftherios Fountoukidis, Ioannis Hadjigeorgiou, Thomai Karamitsou, Anastasios Lytos	SID

Internal Review History

Name	Institution	Date
Grigorios Koutantos	PPC	May 29, 2023
Athanasios Tziouvaras	Bi2S	May 29, 2023

Quality Manager Revision

Name	Institution	Date
Dimosthenis Ioannidis	CERTH	May 29, 2023

Legal Notice

The information in this document is subject to change without notice.

The Members of the EVIDENT Consortium make no warranty of any kind about this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose.

The Members of the EVIDENT Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental, or consequential damages in connection with the furnishing, performance, or use of this material.

The European Commission is not responsible for any use that may be made of the information it contains.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Explanation
ACID	Atomicity, Consistency, Isolation, and Durability
ANSI	American National Standards Institute
APB	Average Price Bias
API	Application Programming Interface
AWS	Amazon Web Services
BI	Business Intelligence
CD	Continuous Development
CI	Continuous Integration
CRM	Customer Relationship Manager
DCE	Discrete Choice Experiment
DOI	Digital Object Identifier
DRY	Don't Repeat Yourself
DX.X	Deliverable X.X
ERD	Entity Relationship Diagram
EU	European Union
GCP	Google Cloud Platform
HPE	Hewlett Packard Enterprise
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
IOS	iPhone operating system
IoT	Internet of Things
IT	Information Technology
JSON	JavaScript Object Notation
ML	Machine Learning
MVC	Model View Controller
MVT	Model-View-View
OLAP	OnLine Analytical Processing
ORDBMS	Object-Relational Database Management System
RBAC	Role-Based Access Control
RCT	Randomised Controlled Trials
RDBMS	Relational Database Management System
REST	REpresentational State Transfer
SQL	Structured Query Language
UI	User Interface
URL	Uniform Resource Locator
UX	User Experience
VPC	Virtual Private Cloud

WP	Work Package
XML	eXtensible Markup Language

Executive Summary

This deliverable focuses on the datahub services of the EVIDENT platform which is developed in the framework of Work Package (WP) 6 "System Integration, Verification and Validation". The deliverable presents the integrated datahub component's specifications, utilities and performance attributes.

Previous and ongoing research has presented significant proof that financial literacy influences household spending and saving decisions, ultimately impacting an individual's financial well-being [1]. The European Union (EU) is host to both the top achievers in the world in terms of financial literacy (Denmark, Sweden), as well as worse performers (Portugal, Romania) that perform below global standards [2]. The EU member states are committed to a 2030 framework for energy and climate change, which establishes novel and demanding objectives for the post-2020 carbon-neutral framework. EVIDENT aligns with the guidelines and frameworks set forth by the EU, employing a comprehensive approach to explore the underlying mechanisms of consumer behavior by leveraging a diverse range of research methodologies, comprising randomised controlled trials (RCTs), experiments, surveys, and case studies. These research methods are complemented by advanced econometric techniques and big-data analytics, which enable the identification of key factors that influence consumer decision-making.

To that end, the project's platform develops and utilises a hub to store all the controlled trial data (respecting individual privacy), econometric analysis results created over the project's duration, and reports and conclusions produced from the experiments and findings.

1. Introduction

1.1 Purpose of the Deliverable

This document is the technical deliverable D6.4 "Datahub services of EVIDENT platform" originating from Task 6.2 "System Development". This deliverable aims to analyze technical advancements of the EVIDENT platform, focusing the integration of the datahub component. Specifically, it describes the development approach used to create the datahub component, which is based on the Continuous Integration and Continuous Deployment (CI/CD) paradigm. The deliverable provides an insightful analysis of the new platform features and demonstrates how numerous functional and non-functional requirements are successfully met in the datahub component. Additionally, it provides a thorough explanation of the technical features, highlighting the improvements and demonstrating how the datahub component was seamlessly integrated.

1.2 Relation with other Deliverables and Tasks

Task 6.2 obtains input from the platform requirements derived from preceding work packages (WPs) in the EVIDENT project, including "WP1 - Requirements and nature of behavioural biases," "WP2 - Policy interventions and pilots design," "WP3 - Intervention, preparation, and execution," and "WP4 - Econometric analysis and policy evaluation." The "System Architecture and Design Specifications" deliverable, which is the primary input document, compiles the specifications from the earlier WPs and serves as the foundation for the system architecture and design specifications. The system developed in Task 6.2 forms the core software foundation for Deliverable 6.3, "Gamification Tools of EVIDENT," and Deliverable 6.4, "Datahub services of EVIDENT platform." Finally, Task 6.3, "System Integration, Verification and Validation," utilizes the platform for the integration and testing plan, culminating in Deliverable 6.5, the "Verification and Validation Report" for the final version of the EVIDENT platform.

1.3 Structure of the Document

The structure and composition of this document have been carefully crafted to deliver a comprehensive analysis of the current state of datahub services, offer clarifications regarding the datahub services within the EVIDENT platform, present a broad overview of the platform to enhance the contextual understanding of the datahub services, explore practical use cases and applications, and provide detailed technical specifications and user navigation information for the datahub component. Each subsequent section is dedicated to examining specific facets of the subject matter, ensuring a cohesive and detailed exploration of the topic:

- Section 2: This examination provides a comparative analysis and understanding, highlighting the strengths and weaknesses of different platforms and setting the stage for the subsequent exploration of the EVIDENT platform's datahub services.
- Section 3: In this section, the overview of the EVIDENT platform sets the foundation by providing a comprehensive understanding of its features, functionalities, and intended applications. Additionally, the unique capabilities and key advantages of the EVIDENT platform are explored,

emphasizing its distinguishing features and potential benefits for users. Furthermore, the analysis of the datahub services delves into the specific functionalities and advantages of this component, showcasing its role in facilitating data management and collaboration capabilities. Lastly, the section covers user roles and terminology, defining the various roles within the platform and elucidating their respective responsibilities and privileges.

- Section 4: The long-term vision of the platform is described here, showing the aim towards building a diverse and inclusive community for conducting free online lab experiments. The benefits of removing the financial restrictions and working with individuals with different backgrounds are showcased, rendering the EVIDENT platform always relevant and ensuring meaningful and inclusive research outcomes.
- Section 5: It provides a user-centric perspective, guiding users through the navigation and functionalities of the Datahub component, enabling efficient and effective utilization of its services. Additionally, the in-depth exploration of the Datahub component provides a comprehensive understanding of its architecture, features, and capabilities, ensuring users clearly understand its role and potential benefits within the EVIDENT platform.
- Section 6: This section concludes the deliverable

2. Similar Platform Features and Best Practices

Data warehousing systems serve as large-scale repositories, software suites and digital infrastructures that store, manage, and analyse vast volumes of structured and semi-structured information from a variety of sources. Such tools are frequently used by companies to increase the efficiency of their decision-making process by assisting with trend analysis, forecasting, and strategic decision-making. For this reason, they are considered the foundation for business intelligence (BI) and data analysis, allowing firms to derive important insights from huge data sets. Their major advantage is that they provide data querying and reporting features by utilizing technologies such as data unification, storage, model creation, management, protection, query and reporting tools. To achieve this, they combine and translate data from many sources, such as operational databases, customer relationship management (CRM) systems, and Internet of Things (IoT) devices.

Software for data warehousing frequently implements measures for data protection and privacy. This trend has been on the rise in recent years due to the introduction of data protection and compliance regulations both at the EU level and within individual countries. In the following subsections, the consortium offers a summary of the most commonly used data warehouse solutions, elaborating on their characteristics, potential drawbacks, and cost specifics.

2.1 Amazon Redshift

Overview: Amazon Web Services (AWS) provides Amazon Redshift [3], a cloud-hosted data warehousing solution. Redshift is designed to be both scalable and cost-effective and as a result, it can handle large datasets and execute queries quickly, making it suited for both SMEs and large corporations. Given AWS's massive resources and infrastructure, Redshift is an excellent solution for enterprises looking to transition away from in-house data warehouses or those just starting out. Figure 1 illustrates the Amazon Redshift user interface.

Features: Amazon Redshift employs columnar storage technology which significantly improves query speed and reduces the volume of data accessed during query operations. Furthermore, Redshift uses clever compression algorithms to reduce storage requirements, increasing speed and economy. This platform seamlessly integrates with other AWS services, such as S3 for data preservation and AWS Glue for data unification, making it easier to create a full data pipeline. Redshift supports SQL and is compatible with a variety of BI and analytics applications, providing flexibility and ease for data analysis. It also includes strong security features, such as data encryption and network isolation through a Virtual Private Cloud (VPC), which ensure that private information is protected and compliance standards are met.

Cons: One area where Redshift might improve is its usability, since some customers report a high learning curve. However, copious documentation and active community participation assist in overcoming this problem. Another issue is that Redshift is a proprietary solution, which means clients are locked into the AWS environment and may encounter difficulties switching to other platforms. Furthermore, despite the fact that Redshift is built to manage enormous workloads, some customers have noticed decreased query performance during peak hours or with particularly big datasets.

Pricing: Amazon Redshift has a pay-as-you-go pricing model, which means that enterprises only pay for the resources they utilize. Pricing is determined by the instance type and region selected, as well as the

quantity of storage and backup required. In this pricing model, it is also possible to reserve instances for a longer period of time, which can result in cost savings over time.
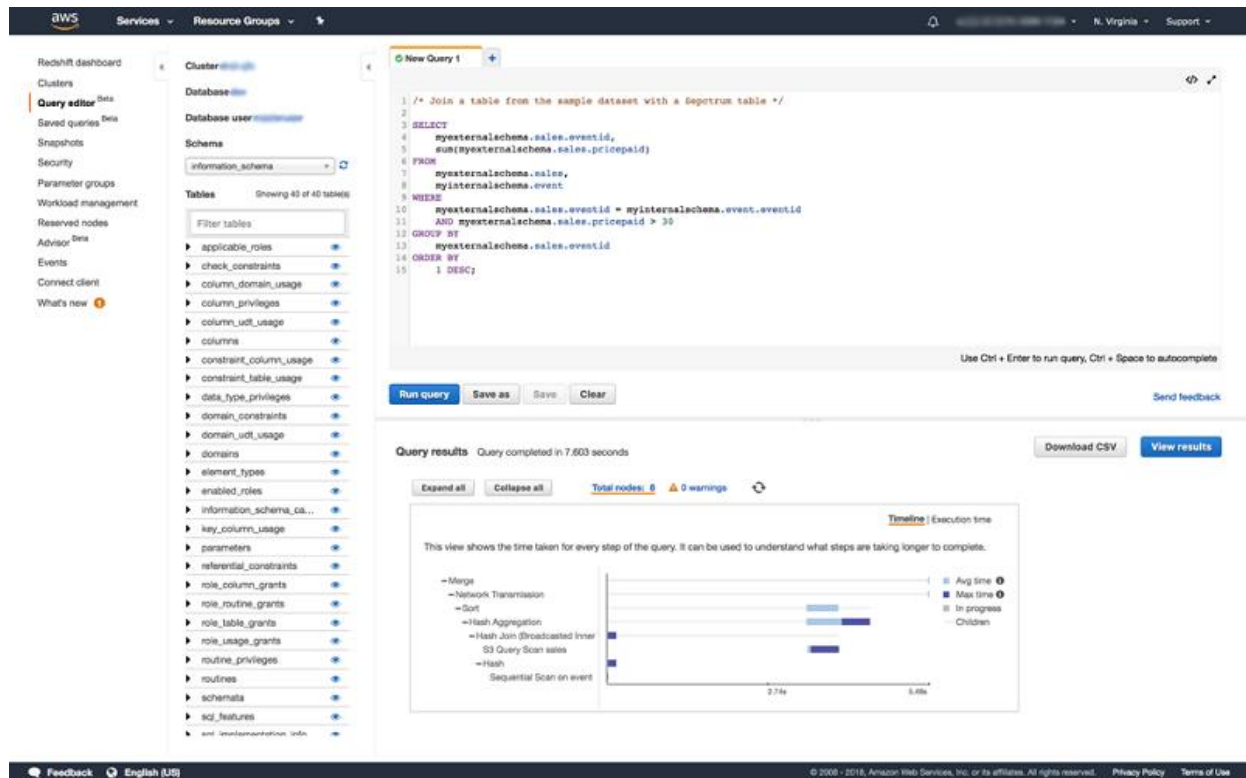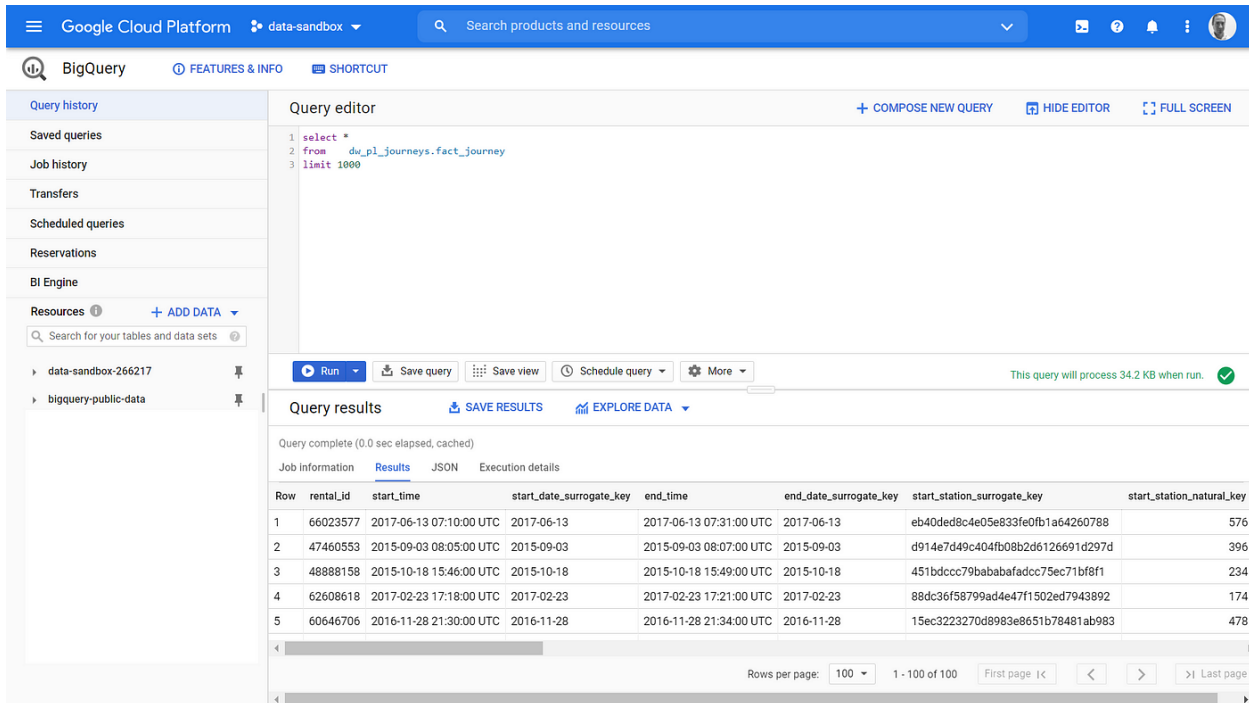

Figure 1: Amazon Redshift user interface

2.2 Google BigQuery

Overview: Google BigQuery [4] is a Google Cloud Platform (GCP) serverless data warehousing solution. It is built to manage enormous amounts of structured and semi-structured data, offering real-time insights to businesses of all sizes. Figure 2 depicts the Google BigQuery user interface.

Features: BigQuery uses a serverless design, which eliminates the need to manage the underlying infrastructure and enables smooth resource scalability. This guarantees that enterprises can handle variable workloads efficiently without having to worry about provisioning or capacity planning. BigQuery also supports ANSI SQL and works well with GCP's machine learning and artificial intelligence services, allowing customers to do sophisticated analyses. BigQuery's interoperability with popular BI tools, as well as its ability to handle and analyze data from a variety of sources, including real-time streaming data, expands its data analysis capabilities. Furthermore, BigQuery offers robust security features like data encryption at rest and in transit, identity and access control, and private networking solutions, ensuring that sensitive data is safeguarded and in compliance with legal requirements.

Cons: BigQuery employs a complicated pricing strategy, which can make cost estimates and management difficult. Furthermore, while BigQuery is intended for high-performance analytics, performance concerns have been noticed, particularly when firms work with very big datasets or complicated queries that are not optimized for the platform.

Pricing: In terms of price, Google BigQuery has a consumption-based pricing model that charges customers depending on the quantity of data processed during searches and the amount of storage used. Long-term storage and real-time streaming data inputs are charged separately. Although BigQuery has a free tier with limited capacity, prices for enterprises with significant workloads or big volumes of data can quickly grow.

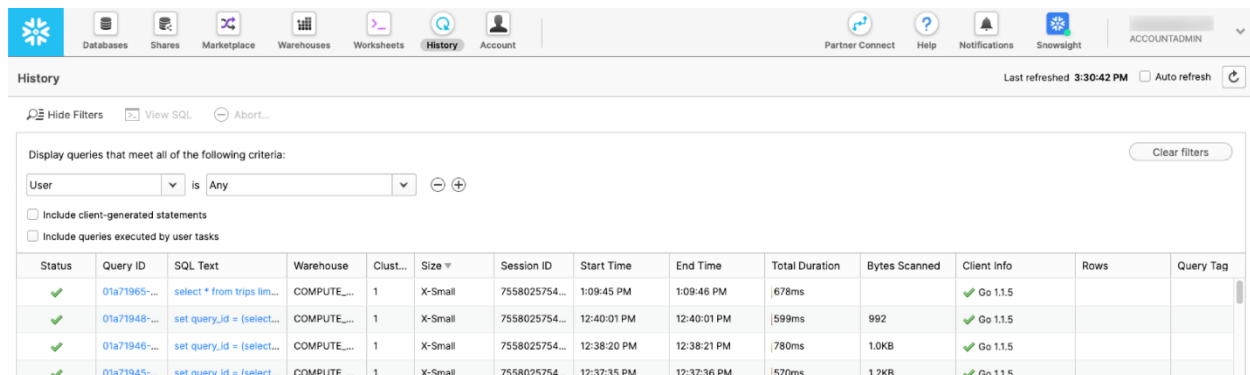Figure 2: Google BigQuery user interface

2.3 Snowflake

Overview: Snowflake [5] is a cloud-based data warehouse solution that has grown in popularity due to its good scaling ability. This solution offers platform-independent features and thus, it may be installed on a variety of cloud providers and cloud infrastructures. Figure 3 illustrates Snowflake's user interface.

Features: Snowflake adopts a multi-cluster design which treats storage and computing resources as different assets. This enables users to expand their data storage and processing capacities independently, resulting in increased performance and better cost efficiency. An important feature of the platform is the support for SQL and its interconnectivity capacity with a wide range of BI tools. This solution also emphasizes security and compliance mechanisms by deploying features such as data encryption, role-based access control, and audit trails to ensure that sensitive information is safely guarded.

Cons: Snowflake's pricing model is consumption-based, with separate charges for storage and computing resources. This may cause costs to easily add up for businesses with huge databases or high processing requirements.

Pricing: Snowflake has a pay-as-you-go pricing model, with separate payments for data storage and computation resources, which are invoiced per second.

Figure 3: Snowflake's user interface

2.4 Microsoft Azure Synapse Analytics

Overview: Microsoft Azure Synapse Analytics [6] is a holistic cloud-based data warehousing and analytics service that amalgamates big data with data warehousing capabilities. As part of the Azure ecosystem, this integrated solution becomes a compelling choice for a wide range of users. The user interface of Microsoft Azure Synapse Analytics is shown in Figure 4.

Features: Azure Synapse Analytics combines big data and data warehousing into a single service, simplifying data intake, preparation, administration, and service for real-time BI and machine learning (ML) applications. Its interoperability with other Azure services creates a unified analytics solution that refines the data flow and improves data analysis capabilities. Along with its data security features, Azure Synapse Analytics provides both allocated and serverless resources, allowing enterprises to select the best option for their unique needs and budget requirements.

Cons: The close linkages between Azure Synapse Analytics and the Azure ecosystem may result in vendor lock-in, complicating the process for enterprises looking to switch to a new platform in the future. Furthermore, the platform's complexity may result in a large learning curve for personnel who are new to data warehousing and analytics, which may impair adoption rates and efficiency.

Pricing: Regarding cost, Azure Synapse Analytics utilizes a pay-as-you-go pricing structure which separates charges for data storage and processing. The cost associated with allocated resources depends on the chosen performance tier, whereas serverless resources are charged on a per-query basis.
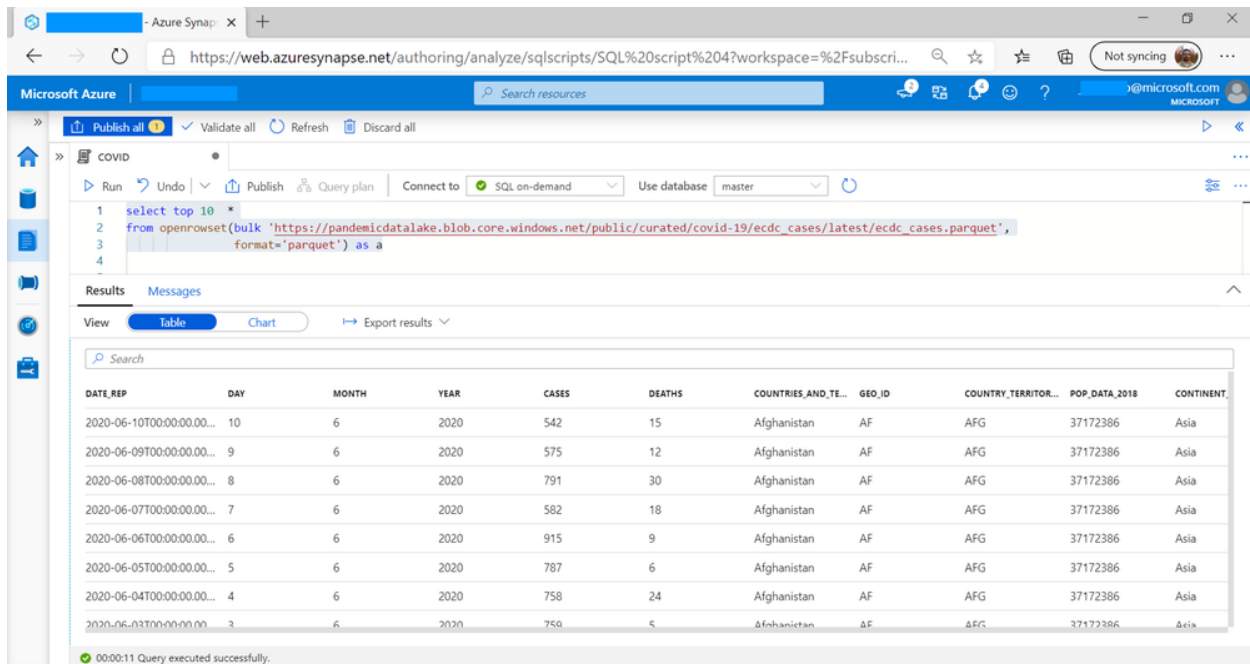
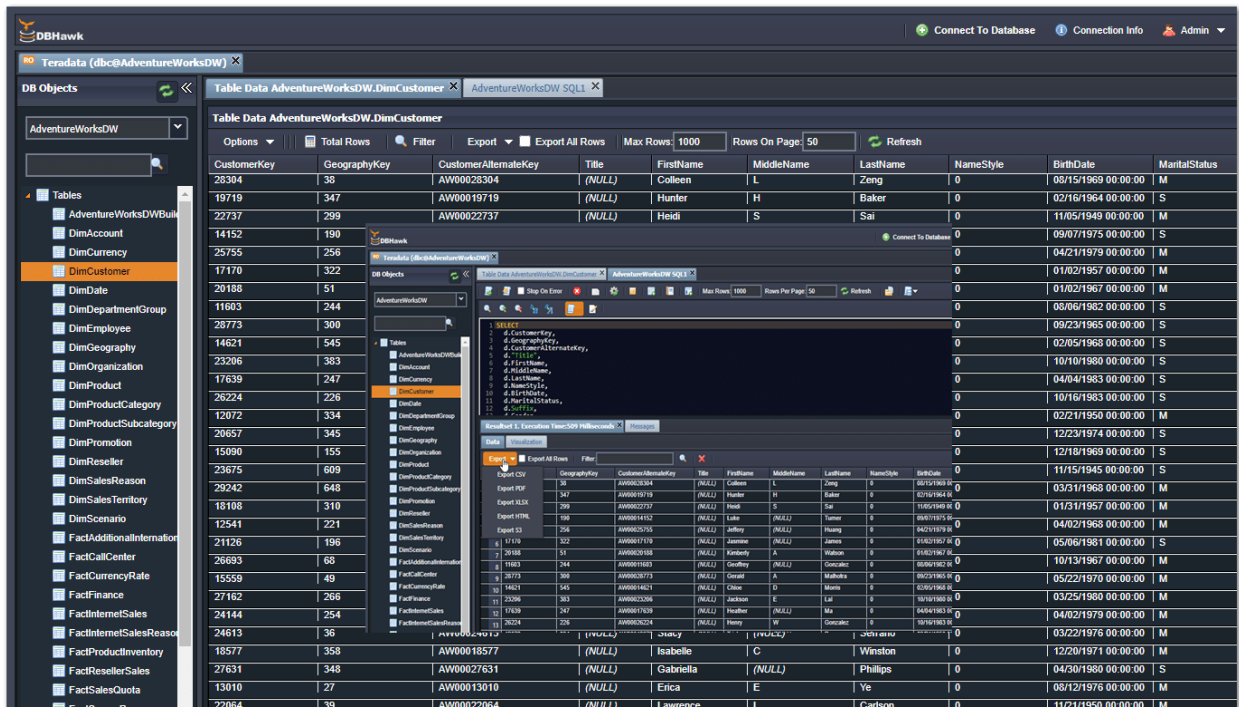**Figure 4: Microsoft Azure Synapse Analytics' user interface**

2.5 Teradata

Overview: Teradata [7] is a well-known data warehousing and analytics system which is offered in on-premises, cloud-based, and hybrid deployment options, making it suited for enterprises with a wide range of requirements and infrastructure choices. The Teradata user interface is illustrated in Figure 5.

Features: One of the key assets of Teradata is its high-performance analytics engine, which has been developed to handle complicated queries and large-scale data processing effectively. The platform is versatile and supports a wide range of data formats, while also offering extensive integration features, allowing it to easily integrate with a variety of BI tools. Teradata offers strong security features, to safeguard user privacy and reduce the chance of data leakage. It also supports cross-platform analytics, enabling users to ingest and evaluate data from other platforms and data warehouses.

Cons: One of the major drawbacks of the platform is its price since Teradata is more expensive than some of the newer cloud data warehousing options. As a result, it may be less appealing to SMEs or companies with restricted expenditures. Another drawback is its technological stack, which may be viewed as less nimble and adaptive in comparison to some of its competitors, especially when serverless is considered.

Pricing: Teradata's price model is defined by two factors: The first one is the deployment option selected by the users, and the other one is the individual product edition. Teradata offers consumption-based pricing for cloud-based installations, with separate costs for data storage and computation resources. Costs also differentiate between on-premises and hybrid implementations which are both supported by the platform.

Figure 5: Teradata's user interface

2.6 IBM Db2 Warehouse

Overview: IBM Db2 Warehouse [8] is an adaptable and high-performance data warehousing solution designed to meet the data demands of enterprises in a variety of sectors. As part of the IBM Db2 family solutions, it leverages IBM's data management and analytics tools to manage vast volumes of information. The user interface of the IBM Db2 Warehouse is shown in Figure 6.

Features: The data analytics features of IBM Db2 Warehouse include built-in support for ML, geospatial analytics and advanced SQL operations. The platform's interoperability with a varied set of data types and formats offers flexibility in dealing with disparate data sources. IBM Db2 Warehouse is also designed with performance-enhancing characteristics such as in-database analytics, in-memory processing and data skipping. Another important feature of IBM Db2 Warehouse is the support of multi-cloud and hybrid data integration, which enables enterprises to combine data from many sources. Being in line with other data warehouse solutions, Db2 Warehouse interfaces with a variety of data integration tools and BI applications, easing the data flow and the analysis process. The architecture of the solution is also built to accommodate a range of deployment choices, including on-premises, cloud and hybrid environments, while also supporting a set of security capabilities. In terms of resource scaling, Db2 Warehouse is a very flexible platform since it provides both scale-up and scale-out features to enable users to optimize resource utilization according to their individual requirements.

Cons: Complexity and cost are the most prominent IBM Db2 Warehouse's disadvantages. The platform's high complexity can be difficult for users inexperienced with data warehousing solutions, while the cost of the solution tends to be higher compared with similar solutions.

Pricing: Depending on the deployment and product edition, IBM Db2 Warehouse provides a number of license options. For cloud installations, IBM uses a consumption-based pricing model with separate costs for storage and computation resources. On-premises and hybrid implementations often include upfront hardware and software expenditures, as well as ongoing maintenance and support fees.
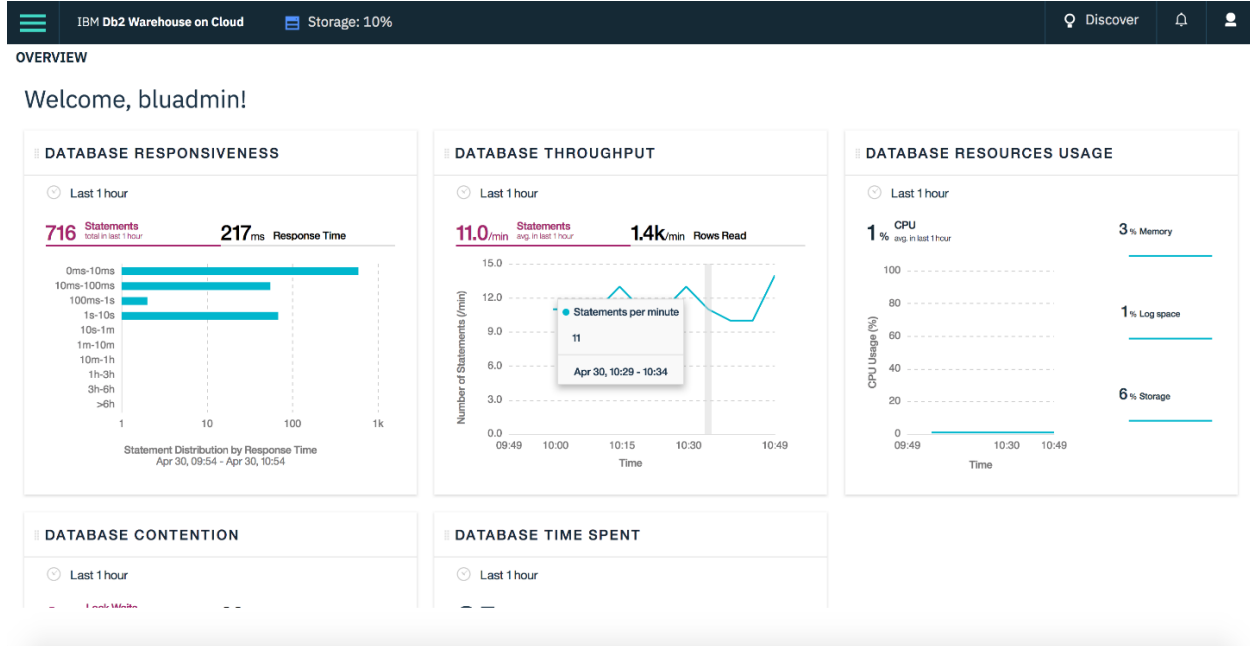
Figure 6: IBM Db2 Warehouse's user interface

2.7   Oracle Autonomous Data Warehouse

Overview: Oracle Autonomous Data Warehouse [9] is a cloud-based data warehousing system that uses AI and machine learning to automate many functions of data administration and optimization. It provides several features to enterprises who already use Oracle products or to those looking for a warehousing solution as part of the Oracle Cloud ecosystem. The user interface of the Oracle Autonomous Data Warehouse is depicted in Figure 7.

Features: Oracle Autonomous Data Warehouse differentiate from its competitors by supporting self-driving features under which no need for human intervention is required. Such features include autonomous provisioning, patching, and performance optimization. The platform employs analytics features, such as ML support and interaction with other BI tools, empowering users to extend their operations to other tools as well. Oracle Autonomous Data Warehouse also includes data encryption, automated backups, and fine-grained access restrictions for further protection and privacy.

Cons: The platform's high cost in comparison to some of its competitors is one possible issue. Another disadvantage of the platform is the inability to use numerous Oracle technologies, including Database Vault, OLAP, Spatial, Text, and Workspace Manager.

Pricing: Oracle Autonomous Data Warehouse utilises a pay-as-you-go pricing approach. The platform provides several price tiers depending on performance levels and storage capacity and thus, enterprises have a lot of pricing options to go with.
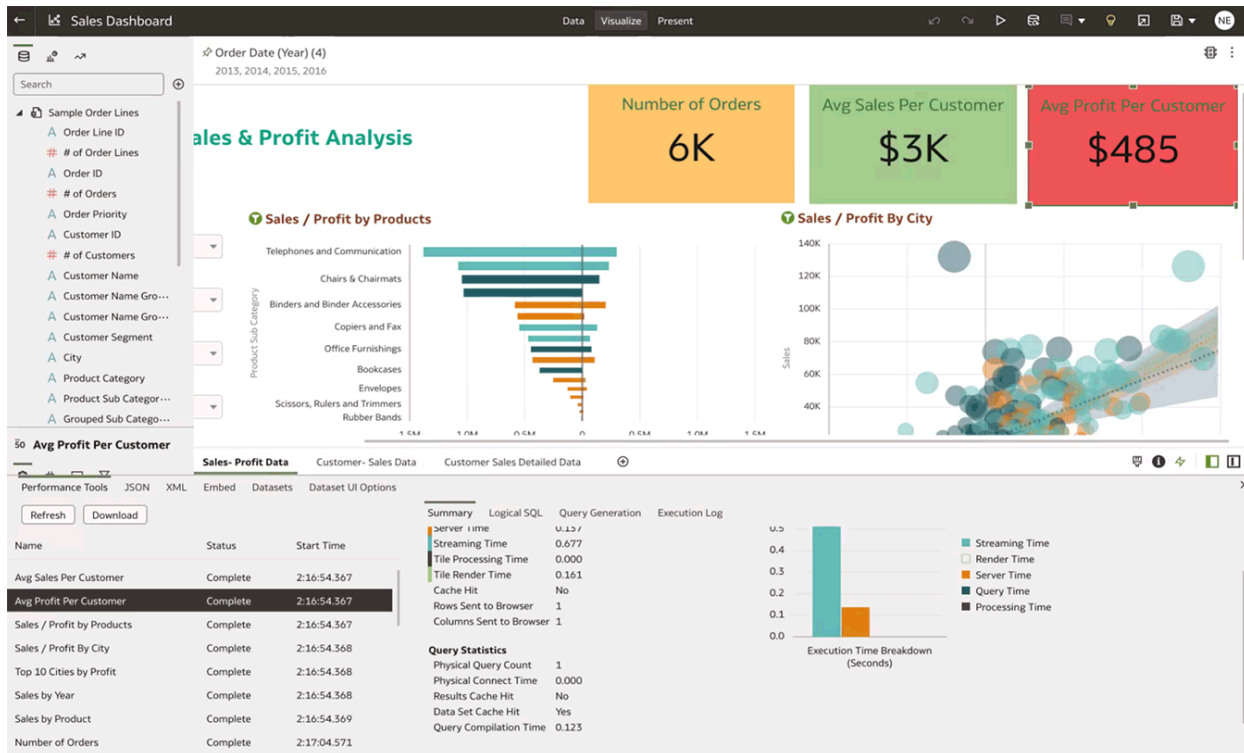
Figure 7: Oracle Autonomous Data Warehouse's user interface

2.8 Databricks Lakehouse Platform

Overview: Databricks Lakehouse Platform [10] is an appealing alternative for enterprises wishing to embrace a contemporary data architecture since it allows easy connection with major cloud providers. Figure 8 shows the Databricks Lakehouse Platform's user interface.

Features: One of the Databricks Lakehouse Platform's key assets is its ability to combine batch and streaming data processing, enabling enterprises to execute near-real-time analytics fast and in a more efficient way. The platform makes use of the open-source Delta Lake storage layer, which improves data dependability and speed by including capabilities like ACID transactions and schema enforcement rules. Databricks supports popular programming languages, such as Python, R, and Scala, and integrates well with other cloud solutions. Finally, the platform provides security and compliance features, to assist enterprises in protecting sensitive data while remaining compliant with industry regulations.

Cons: The platform is built on the Apache Spark ecosystem, which may impose a learning curve for individuals unfamiliar with Spark or other data processing frameworks.

Pricing: The Databricks Lakehouse pricing model is based on usage. Similar with other data warehouse solutions, the platform provides several price levels based on desired performance and features.

Figure 8: Databricks Lakehouse Platform's user interface

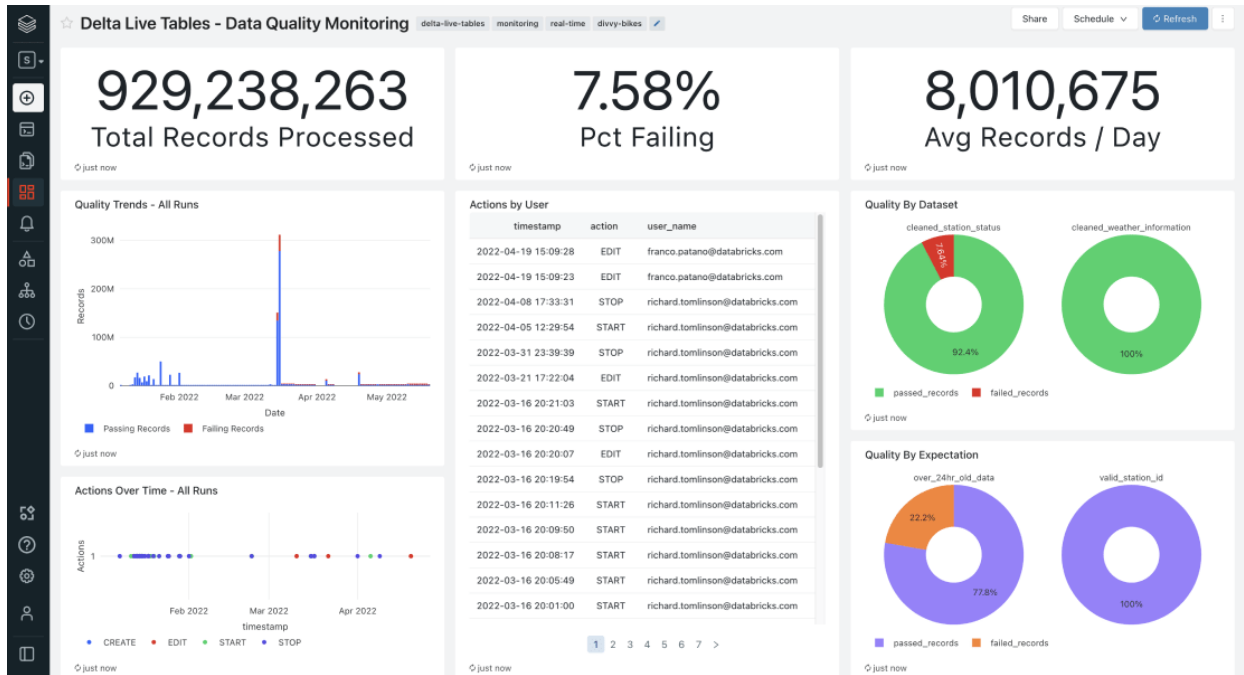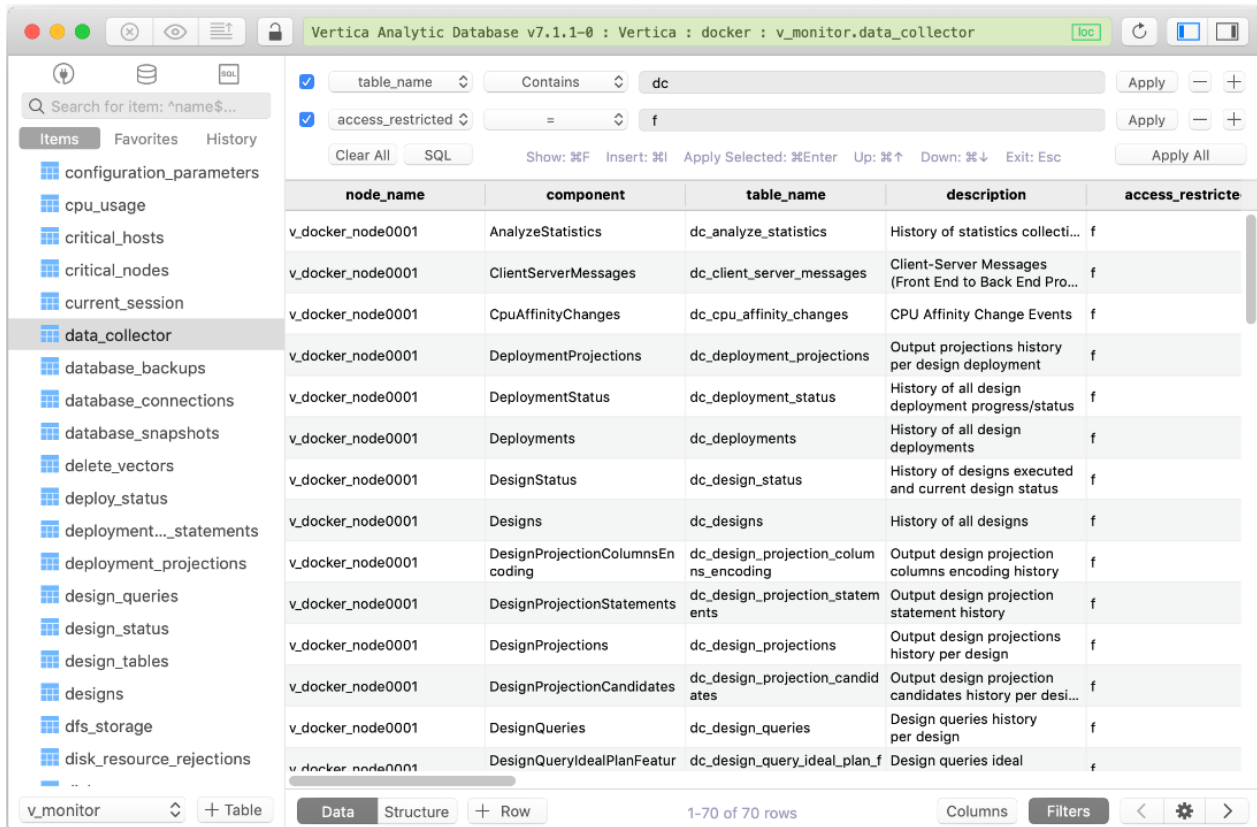## 2.9 Vertica

Overview: Vertica [11] is a high-performance and scalable data management platform built to handle large amounts of data and support analytics workloads. The platform is developed by Vertica Systems and is offered in a variety of deployment choices, including on-premises, cloud, and hybrid environments. Figure 9 illustrates the graphical user interface of the Vertical platform.

Features: Vertica's columnar storage design provides data compression and high-speed query performance, making it particularly ideal for large-scale data analytics. The platform also has extensive analytics capabilities and interfaces with common programming languages, such as Python, R, and Java. Vertica prioritizes scalability and performance, enabling both horizontal and vertical scaling to meet fluctuating workloads and large data quantities. Workload management and resource sharing are also available on the platform, allowing users to optimize resource allocation and task scheduling. From the side of security, Vertica provides data encryption, full audit logging and access control to protect sensitive data.

Cons: The platform's drawbacks include the relatively poor design of geographic functions (spatial data processing interfaces), demanding maintenance, and a lack of support for interfaces with other databases.

Pricing: Vertica's pricing structure changes depending on the deployment method and product version chosen. Vertica employs a usage-based pricing method for cloud installations, while for on-premises and hybrid deployment solutions, recurring payments are foreseen. Such payments include initial hardware and software costs and expenses for maintenance and support services.

Figure 9: Vertica Platform's user interface

2.10 Panoply

Overview: Panoply Cloud Data Warehouse [12] is an all-in-one data management platform that provides a solution that serves enterprises of all sizes and sectors, with a heavy emphasis on ease of use and quick implementation. The main focuses of Panoply are its user-friendly interface and its smooth learning curve. Figure 10 illustrates Panoply's user interface.

Features: Panoply widely supports data source integration, allowing users to easily aggregate data from different sources such as databases, APIs, and third-party applications. The platform also has built-in data transformation and enrichment functions, which act as data pre-processing operations by standardizing and cleaning the data. This process can prepare the data for analysis without the need for complex code or third-party applications. Panoply provides interoperability features since it works well with common BI, data warehousing, and analytics solutions. In terms of performance and scaling, the solution supports resource elasticity functions that enable real-time adaptations to the storage and processing capacity.

Cons: One disadvantage of the platform is that its analytical capabilities are less sophisticated when compared to some of its competitors. Furthermore, while Panoply's emphasis on usability is typically advantageous, users may find the platform's customization choices restricted.

Pricing: Panoply's pricing strategy is consumption-based, with several price tiers geared to various usage levels and features.
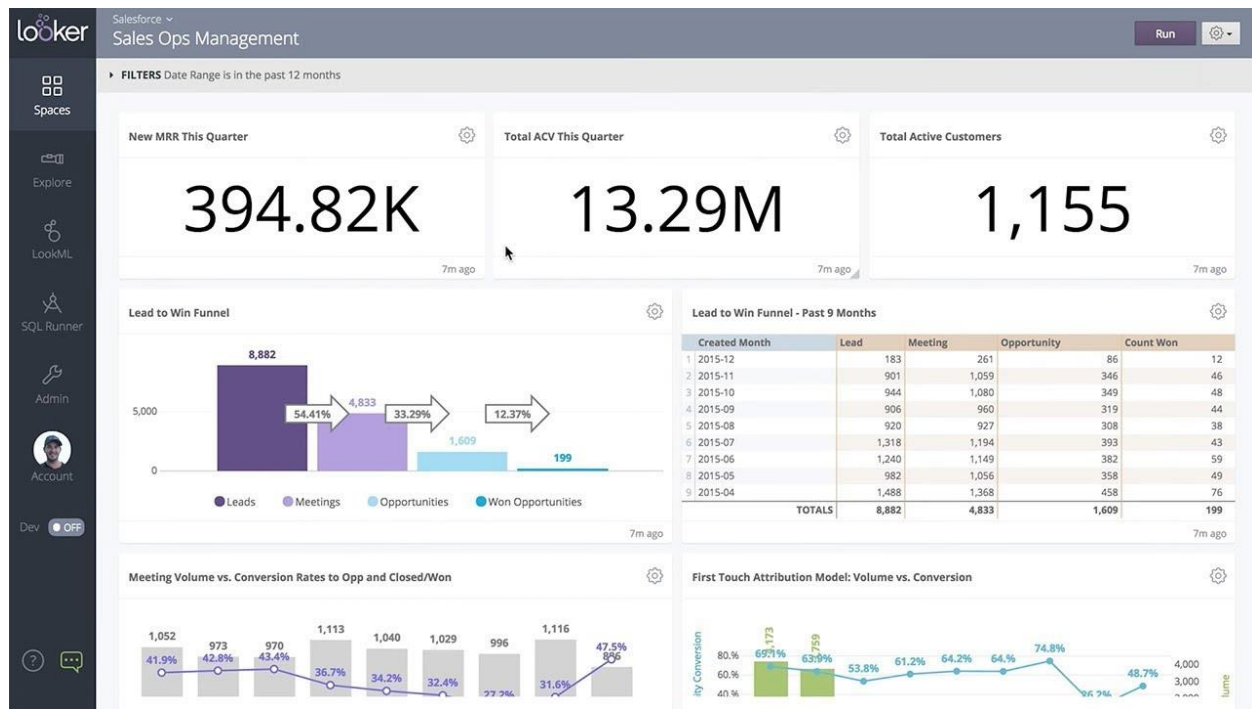
Figure 10: Panoply Platform's user interface

Table 1 depicts the key characteristics of the Data Warehouse Solutions which are discussed in this section.

Table 1: The features, drawbacks, pricing options and market share of the most prominent Data Warehouse Solutions

Solution	Features	Cons	Pricing	Market share
Amazon Redshift	Storage optmisation, Improved query performance, interoperability, robust analytics.	Compatible with AWS only, steep learning curve, minor performance issues.	Pay-as-you-go	29.51% [13]
Google BigQuery	Serverless architecture, ANSI SQL and machine learning support, compatibility with external tools.	Complex pricing model, performances issues when large volume of data is considered.	Consumption-based pricing	14.04% [14]
Snowflake	Multi-cluster architecture, data encryption, good scaling capacities.	Expensive when large amounts of data are involved, not so optimized for data analytics.	Pay-as-you-go	19% [15]
Microsoft Azure Synapse Analytics	Big data support, machine learning capabilities, serverless and provisioned architectures.	High complexity and steep learning curve, difficult to migrate to other platforms.	Pay-as-you-go	10.75% [16]
Teradata	High-performance analytics, interoperability, compatibility with other solutions.	High costs, less agile compared to similar cloud solutions.	Consumption-based and licensing	16.98% [17]
IBM Db2 Warehouse	Built-in support for machine learning, platform compatibility with data types, in-memory processing capabilities, multiple programming language support.	High platform complexity, high cost.	Consumption-based and licensing	5.06% [18]
Oracle Autonomous Data Warehouse	Self-driving capabilities, automatic provisioning, automatic performance tuning, advanced analytics, integration capacity with other tools.	High costs, not compatible with other Oracle features such as: Database Vault, OLAP, Spatial, Text, and Workspace Manager.	Pay-as-you-go	0.29% [19]
Databricks Lakehouse Platform	Streaming data processing, real-time analytics, data Lake support, multiple programming language support.	Relatively high cost, strong reliance in Apache Spark ecosystem which lowers the versatility of the platform.	Consumption-based	N/A
Vertica	Efficient compression, high performance, integration with popular programming languages, high scalability.	Poor maintainability, lack of integration with other databases.	Consumption-based and licensing	0.29% [20]
Panoply	Easy-to-user, interoperability with other platforms, good scalability, resource optimisation capabilities,	Not so good analytics engine, limited customization options.	Consumption-based	0.24% [21]

3. Overview of the EVIDENT Platform

This section provides a comprehensive overview of the EVIDENT platform, by also analysing the core functionalities and features. Specifically, it presents the platform's architecture, while also describing the design of the various architectural components. Additionally, it highlights the capabilities and significant benefits of the EVIDENT platform, underscoring its unique value proposition. Furthermore, this section provides an in-depth explanation of the datahub service incorporated into the EVIDENT platform. Finally, the user roles and terminologies employed within the platform are introduced. By offering an enhanced and detailed account, this section aims to provide a comprehensive understanding of the EVIDENT platform and its notable features, giving the necessary context to the reader for the upcoming datahub specifications of this deliverable.

The EVIDENT platform encompasses a unified endpoint that integrates various components including crowdsourcing tools, serious game applications, and datahub services. A pivotal element of this platform is the datahub service, which plays a vital role in facilitating seamless storage, systematic organization, and swift retrieval of experiment-generated data. The datahub service ensures the integrity, accessibility, and optimal utilization of valuable data derived from the experiments conducted within the platform. By harnessing the collective intelligence and diverse expertise of a crowd, the platform aims to enhance the accuracy, efficiency, and quality of its experimental endeavours.

In the following sub-sections, the intricate architecture of the EVIDENT platform (section 3.1) will be examined, encompassing its underlying framework and crucial components. Moreover, the unique characteristics that render the EVIDENT platform one of the best and most inclusive solutions on the market are elaborated in section 3.2. The focus will extend to the comprehensive exploration of the Datahub Services (section 3.3) provided by the platform, specifically concentrating on their advanced functionalities for robust storage, meticulous organization, and expedient retrieval of experiment-generated data. Moreover, we will meticulously analyse the dynamics of User Roles and Terminology (section 3.4) prevalent within the platform's ecosystem, elucidating the distinct roles assigned to organisers and participants and the precise terminologies employed to foster coherent communication and effective collaboration. With this rigorous investigation, the objective is to provide a better understanding of the EVIDENT platform's sophisticated architecture, considering its cutting-edge data management capabilities via Datahub Services component utilisation.

3.1 General architectural description of EVIDENT platform

The EVIDENT platform architecture is meticulously designed to align with the requirements and specifications of its end users. These user-centred concerns come from the development team's consultations with the end users during Task 6.1, "System Architecture and Design Specifications." The development group decided to use the Continuous Integration Continuous Deployment (CI/CD) methodology, which allows specifications to be gathered concurrently with software development, meeting potential EVIDENT consortium demands.

In order to mitigate potential challenges related to future expansion and enhance the long-term sustainability of the EVIDENT platform beyond the project's duration, a component-based design approach was chosen. By adopting a component-based design, the EVIDENT platform ensures flexibility and adaptability, allowing for easier integration of new features, modules, and functionalities if needed.

This design methodology enables the platform to evolve and scale effectively, accommodating emerging technologies and evolving user requirements. Components are developed as independent entities, allowing easier maintenance, updates, and enhancements, boosting the platform's sustainability by promoting modularity and reusability.

The EVIDENT platform architecture comprises five key components, each playing a crucial role in its overall functionality and providing a comprehensive ecosystem for users:

- The Datahub service, which is the main subject of this deliverable, is an online centralised storage for all the datasets utilised by the EVIDENT project. Moreover, it has the capacity to deposit various information such as data collections, software scripts and computational algorithms which renders this service invaluable for future research endeavours. This service operates independently from the following tools as it enables the storage, preservation, and administration of resources with varying levels of accessibility, including public, private, or restricted modes.

- The user component incorporates a wide array of options and tools accessible to users for interacting with the platform's services. This component has been thoroughly elaborated in D6.2 "Crowdsourcing Tools of EVIDENT platform" (section 4.1)

- The data extraction component analyses the gathered data and generates valuable findings utilized for both business and academic objectives (elaborated extensively in the forthcoming deliverable D6.3 'Gamification Tools of the EVIDENT Platform').

- The data collection component aggregates the data provided by the end users, anonymises the information, and transfers it to the aforementioned data extraction component. The functionality of this component is similar to the session as it involves the framework for organizing and preserving the accumulated data resulting from the users' involvement in an experimental setting. Extensively discussed in D6.4 "Datahub services of the EVIDENT platform" in subsection 5.2 'Additional features' where a detailed description can be found.

- The crowdsourcing component collects feedback from platform users actively engaged in the hosted sessions featuring published surveys and serious games. Both the serious game application and the survey subcomponent are elaborated upon within this deliverable and have been extensively detailed in D6.2 'Crowdsourcing Tools of the EVIDENT platform' (subsection 4.2).

Figure 11 displays a visual representation of the platform's architectural framework, illustrating the interconnections among the various components comprising the platform. The diagram visually depicts the relationships and interactions between the different elements of the platform.
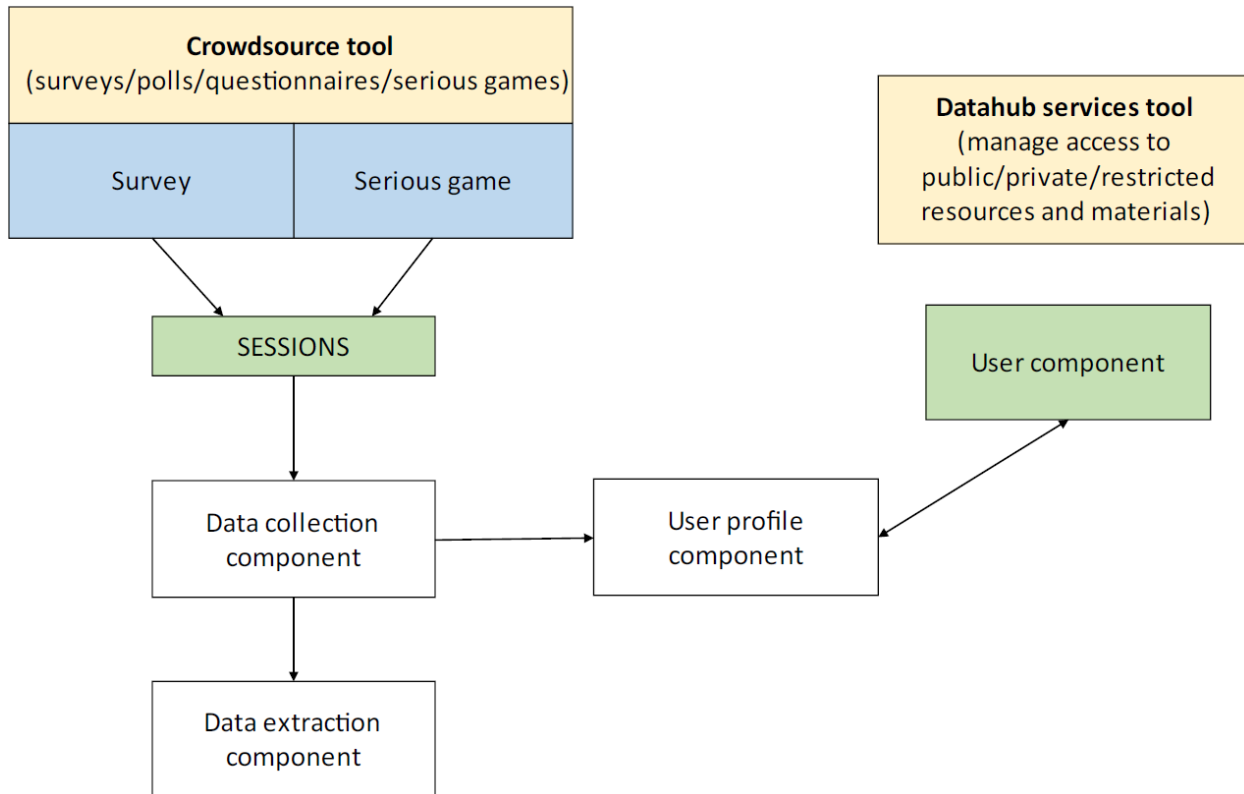
Figure 11: Architecture of the EVIDENT platform components

The session component facilitates the seamless functioning of the core feature of the EVIDENT platform, enabling users to effortlessly join and engage with lab experiments. On the other hand, the user profile component serves as a pivotal connector between the key functionalities of the EVIDENT platform, encompassing activities such as account creation, participation in sessions, integration of serious games, and more, and the datahub services tool.

The crowdsource module, which examines the complex structure of the platform, is intended to get user opinions on the common prejudices that influence their preference for the consumption of energy. Data collected through the survey and serious game sub-components provide knowledge and comprehension of customers' viewpoints and behaviours, bringing insightful knowledge about their thoughts and behaviours. Organisations participating on the platform can create and improve surveys to be used in research studies, by utilising the survey tool. Once the survey has been made public, it is considered another step within the context of a lab experiment. In terms of the survey subcomponent, the EVIDENT framework is making use of SurveyJS, a well-known JavaScript-based survey builder. By integrating with the platform's internal data storage system, this connection intends to offer an advanced and user-friendly survey generation tool. Users inside the organisation can have access to the full set of features provided by the SurveyJS library, allowing them to create interactive surveys and effortlessly include them in their lab studies.

3.2 Unique capabilities and key advantages of EVIDENT platform

The integration of surveys and serious games has a synergistic effect that improves the whole research experience, encompassing the EVIDENT platform's unique value proposition. The platform provides an innovative and complete solution for running lab experiments, as well as acquiring important data by combining these two strategies.

The EVIDENT surveys offer stakeholders a way to obtain specific data from respondents, because of their organised and methodical character. They provide greater comprehension of participants' perspectives, beliefs, and actions in relation to the study topics, by enabling researchers to collect both quantitative and qualitative data. As a standardised approach, surveys are incorporated into EVIDENT to make it easier to compare data from different participants and analyse it.

On the other hand, the serious game module of EVIDENT introduces an element of gamification into the research process. These interactive and engaging games serve as an innovative and dynamic medium to capture participants' attention, foster active involvement [22], and elicit behavioural responses. Serious games have the potential to create a more immersive and enjoyable research experience [23], potentially leading to higher participant engagement and more accurate data collection.

All the data generated from the surveys and the serious game are stored in the datahub service of the EVIDENT platform. The datahub service serves as a centralized repository, ensuring the efficient and reliable storage of the collected data. By leveraging this dedicated service, the platform offers a robust infrastructure that guarantees the integrity and accessibility of the experiment-generated data.

An additional noteworthy value offered by the EVIDENT platform is its availability to organizations for conducting e-lab experiments at no cost. The platform's accessibility without financial barriers allows organizations, particularly researchers and academics, to leverage its capabilities and resources without incurring expenses.

The process of providing the EVIDENT platform freely to organizations fosters inclusivity and promotes scientific collaboration on a broader scale. This availability empowers researchers, regardless of their funding constraints, to engage in innovative e-lab experiments and explore research questions within their fields of expertise. It democratizes access to advanced research tools and technologies, levelling the playing field and facilitating equal opportunities for scientific inquiry. The cost-free nature of the platform also encourages greater participation from diverse organizations, leading to a more extensive range of studies and a richer pool of research insights. It nurtures a collaborative ecosystem where organizations can freely collaborate, share knowledge, and collectively contribute to scientific advancements.

To that end, the EVIDENT platform promotes openness, encourages collaboration, and enables a wider community of researchers to benefit from its features and capabilities. This added value enhances the platform's impact, accessibility, and potential for driving transformative research outcomes.

3.3 Datahub Services

This subsection presents a detailed explanation of the generic functionalities of the datahub service that is provided within the EVIDENT platform. It outlines how the datahub services effectively support various essential operations and workflows within the platform. Furthermore, a more thorough navigation of

EVIDENT's datahub service component is presented in "EVIDENT Datahub Services User Navigation" section of this deliverable.

The datahub services components serve as an online centralized hub for hosting all the data utilized in the intervention trials performed in terms of the EVIDENT project. This data is stored in an anonymised format, ensuring the privacy of the user's personal information. The datahub not only houses the intervention trial data but also encompassed the comprehensive collection of econometric analytical tools created throughout the project's duration. Additionally, the platform has the capacity to keep the documentation and discoveries from the trial outcomes, offering a consolidated repository of valuable information. Through this integrated datahub services tool, the EVIDENT platform ensures secure data management, supports advanced analysis, and enables the generation of valuable insights for further research and decision-making processes. Finally, all the aforementioned utilities have been designed considering the nature of the potential future requirements of other projects and initiatives, thus ensuring the sustainability and lifelong relevance of this platform.

This particular platform component serves as a data gateway, offering EVIDENT users access to a wide range of materials (described above) that align with the project's scope. These materials may be openly available or protected, depending on their nature and requirements. While distinct from the previous components, the datahub services tool plays a crucial role in storing, saving, and managing resources that may have varying levels of accessibility, including being open to the public, exclusive to one or more individuals, or classified access. All users that have been logged in will have access to a system resembling Zenodo, ensuring a robust and reliable repository for storing project-related material. Administrative privileges will empower the designated administrator to add, modify, remove, and examine these materials and resources, ensuring smooth and efficient management within the platform.

The Zenodo repository has been chosen as the main data repository to host the data created on the project's platform. Digital object identifiers (DOIs), which Zenodo incorporates, allow for secure preservation and trustworthy referencing, guaranteeing the data's ongoing identity and accessibility. The EVIDENT platform having smoothly integrated Zenodo, enables stakeholders to produce a number of DOIs that correlate to the datahub, in order to give end users improved functional commodities. Additionally, Zenodo provides an application programming interface (API) that enables swift data synchronisation and transfer while facilitating smooth connection with the EVIDENT Platform. A sandbox environment is accessible for testing reasons, offering a regulated area where developers may carry out integration tests throughout the development phase. Developers may use this sandbox environment as a trustworthy and monitored testing space because it is simple to reset or clean at any moment.

3.4   User Roles and Terminology

The goal of this section is to provide a thorough explanation of the various user roles and related terminology used inside the EVIDENT platform. It gives a thorough rundown of every user position, emphasising the duties, rights, and platform contributions that each one has. This section explicitly defines the roles and responsibilities of each user and allows participants to understand their own contributions, thus encouraging productive cooperation inside the platform. Additionally, it helps users get comfortable with the particular vocabulary used by the EVIDENT platform, which promotes clear communication and improves the user experience overall. Participants will obtain a thorough grasp of their duties and how

they may actively contribute to the platform's goals and continuous improvements through this in-depth discussion of user roles and terminology.

The EVIDENT platform has a powerful role-based access control (RBAC) system, where each user's individual position specifies the rights and privileges that are provided to them on the platform. By using a role-based approach, users are guaranteed to have the proper access privileges according to their allocated roles. In addition to analysing platform information and a list of available sessions, registered users have access to a variety of activities depending on their current role. Non-registered users, on the contrary, are restricted to accessing platform data and a list of available sessions. Non-registered users must complete a registration procedure in order to gain access rights inside the platform, that will allow them to actively engage in the sessions that are provided. The EVIDENT platform uses RBAC methodology to manage user access efficiently and make sure that users are given the right degree of rights, based on their roles and responsibilities.

Within the EVIDENT platform, a well-defined structure of four (4) distinct user roles has been established: super administrators, organizations, participants and visitors, who relate to the unregistered users. Each role caters to a unique target group with distinct goals and preferences within the platform.

Registered organizations on the EVIDENT platform usually are comprised of researchers and scholars, who adopt the platform to design and conduct lab experiments aimed at exploring specific research questions. The platform offers a versatile approach, combining surveys and serious games, to meet their specific de mands and goals. Organisations can modify the lab trials in accordance with their research objectives thanks to this hybrid system. One of the EVIDENT platform's notable features is its provision for organizations to configure the serious games within their lab experiments. This configuration capability empowers organizations to receive input from surveys, thereby creating customized and modular scenarios for each of the specific use cases. By considering participants' demographics, preferences, and knowledge, the platform enables organizations to deliver targeted and relevant content, enhancing participant engagement and optimizing research outcomes.

In the EVIDENT platform, the super-admin role is exclusive and not available for registration. Only existing super-administrators have the ability to register and assign this particular user role. The super-admin role is reserved for individuals with high-level administrative privileges and responsibilities, ensuring the proper management and oversight of the platform's operations. These super-administrators possess special permissions to configure system settings, manage user accounts, and maintain the overall functionality of the platform.

Table 2: User Roles and Corresponding Permissions in the EVIDENT Platform

User roles / capabilities / permissions	Super Administrator	Registered Organiser	Registered Participant	Unregistered Participant
Browse the web page to view the static content.	✓	✓	✓	✓
Identification, Login, Logout, and Password Reset	✓	✓	✓	✓

User profile management	✓	✓	✓	
User account removal	✓	✓	✓	
User profile information export	✓	✓	✓	
User demographics update	✓	✓	✓	
Compile / revise / remove survey	✓	✓		
Upload / revise / remove a serious game	✓	✓		
Compile / revise / remove a new lab experiment	✓	✓		
Serious game configuration for lab experiments	✓	✓		
Take part in a lab experiment	✓	✓	✓	
Check the response to a lab experiment	✓		✓	
Export answers from a lab experiment	✓		✓	
Retrieve information collected during a lab experiment	✓	✓		
Share lab experiment	✓	✓	✓	✓
access to all information in its entirety	✓			

In Table 2 above, a concise overview of the actions that can be performed by each user role within the EVIDENT platform is provided. The table outlines the specific capabilities and permissions granted to super administrators, organisations, and registered and unregistered participants. It comprehensively captures the range of actions and functionalities available to each user role, showcasing their distinct roles and responsibilities within the platform. The information presented in Table 6 serves as a valuable reference, enabling users to understand the scope of their access and the actions they are authorized to undertake based on their assigned user role.

The main ribbon of the EVIDENT platform web page comprises several key components that offer intuitive navigation and functionality. The "About" section provides essential information and details about the questionnaire creation process, the integration of the serious games, the study sharing and the data collection. It serves as a resource for users to understand its purpose, features, and benefits. The "Participate" section enables users to actively engage in the platform's experiments and activities, offering a gateway to contribute their knowledge and expertise. The "Available Data" component grants users access to a pool of available data on the platform, facilitating exploration and analysis of the collected information. Lastly, the "Register" feature allows new users to create an account, and update their demographic information granting them the ability to participate in experiments and contribute to the platform's growth. Together, these web page features provide a user-friendly interface, guiding users through various aspects of the EVIDENT platform and empowering them to actively participate in its activities and access valuable data resources. In Figure 12, this ribbon is displayed.
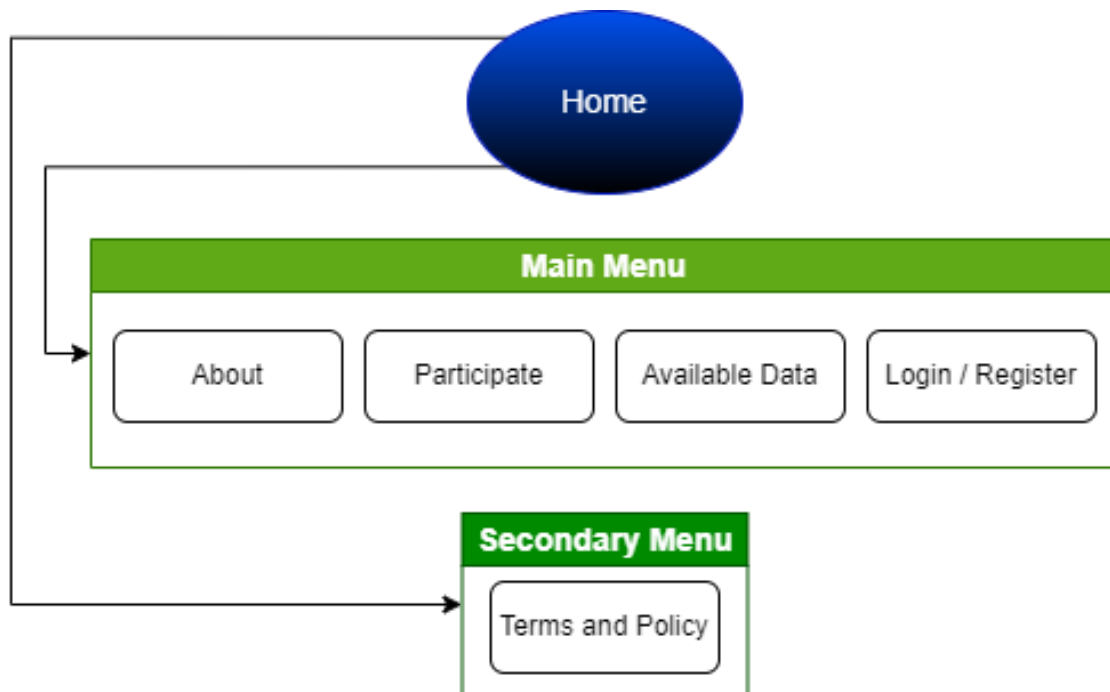
Figure 12: Main ribbon of EVIDENT platform web page

3.5 Updated elements of EVIDENT Platform

Throughout the development cycle of the EVIDENT platform, significant updates and enhancements have been implemented to provide users with a more comprehensive and user-friendly experience. The platform now offers a clear and concise overview of the platform's specifications, ensuring that visitors can quickly understand its features and functionalities. The information provided is presented in a friendly and easy-to-read format, making it accessible to a wide range of users, including researchers, policymakers, community members, and individuals.

One notable module of the platform is the introduction of a newsletter feature. Participants can subscribe to the newsletter and receive regular updates on new sessions and get informed with the latest news. This newsletter serves as a valuable communication tool, keeping participants informed and engaged with the latest developments on the platform and the community.

Figure 13 displays the interface of this module, as shown on the "Home" page of the EVIDENT platform.
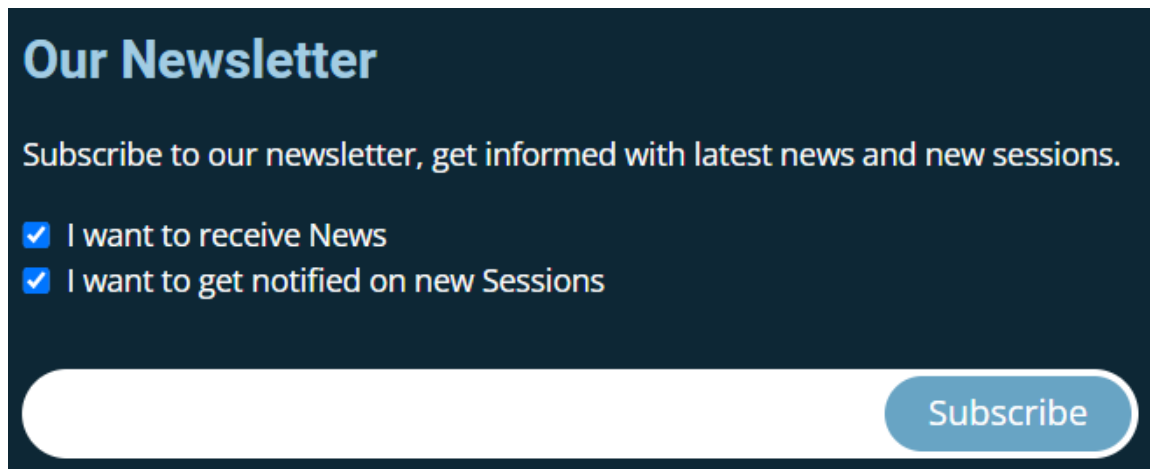
Figure 13: Newsletter subscription at EVIDENT web page

The new enhanced description of the EVIDENT Platform provides a more comprehensive and user-friendly understanding of its capabilities and scope. Figure 14 displays the interactive building blocks of EVIDENT's core components and functionalities in a more intuitive and user-friendly manner. By highlighting features such as questionnaire creation with a wide selection of question types, advanced customization options, and seamless integration with serious games, users can better grasp the platform's versatility and potential for multifaceted research. The emphasis on easy study sharing through various channels, coupled with the utilization of social connectivity, emphasizes the platform's commitment to reaching a wider audience and fostering collaboration. Furthermore, the streamlined data collection process ensures data security and accessibility in various formats, empowering researchers to analyse and derive insights effectively. This improved description enables users to grasp the full potential of the platform, making it an invaluable tool for conducting data-driven research with ease and impact. Finally, Section 0 presents the updated landing pages designed for individuals, researchers and policymakers promoting the unique features of the EVIDENT platform.
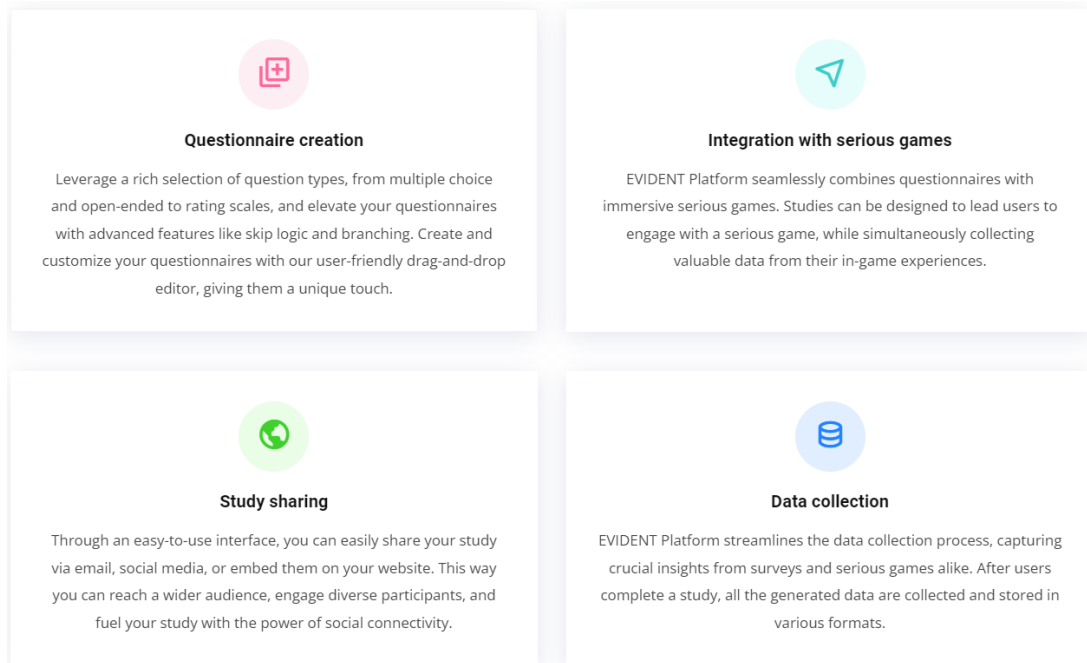
Questionnaire creation

Leverage a rich selection of question types, from multiple choice and open-ended to rating scales, and elevate your questionnaires with advanced features like skip logic and branching. Create and customize your questionnaires with our user-friendly drag-and-drop editor, giving them a unique touch.

Integration with serious games

EVIDENT Platform seamlessly combines questionnaires with immersive serious games. Studies can be designed to lead users to engage with a serious game, while simultaneously collecting valuable data from their in-game experiences.

Study sharing

Through an easy-to-use interface, you can easily share your study via email, social media, or embed them on your website. This way you can reach a wider audience, engage diverse participants, and fuel your study with the power of social connectivity.

Data collection

EVIDENT Platform streamlines the data collection process, capturing crucial insights from surveys and serious games alike. After users complete a study, all the generated data are collected and stored in various formats.

Figure 14: Updated "About" section of EVIDENT web page

4. Various applications to a diverse community

EVIDENT is a flexible platform that provides academics, policymakers, community members, and people with a number of capabilities that enable data-driven research. It gives researchers a strong foundation for interacting with many stakeholders and maximising the effectiveness of their study. Researchers may design complex experiments and link with other researchers, decision-makers, and communities to encourage partnerships and information sharing by using its features and approaches.

While solutions like Amazon Mechanical Turk[1], microworkers[2], and Prolific[3] have become well-liked choices for online data collecting, EVIDENT stands out in a significant way. EVIDENT aims to offer a free alternative to the existing solutions, which demand that researchers pay contributors for their work. EVIDENT intends to democratise access to online lab experiments by removing the price barrier, allowing a wider range of researchers, particularly those with restricted funds, to carry out their investigations and gather high-quality data. This dedication to open access is consistent with EVIDENT's goal of establishing a community of researchers who are inclusive and collaborative and who can share their thoughts and insights without being constrained by financial obligations.

Creating a diverse and inclusive community is an aspirational objective for the EVIDENT Platform. In the long run, this platform aspires to unite people from different backgrounds, allowing them to freely participate in research projects. This dedication to diversity fosters a culture that stimulates communication, encourages teamwork, and raises the standard of research outputs. EVIDENT enhances the research process by bringing in people from many backgrounds, such as researchers, policymakers, community members, and individuals, and promoting novel methods for problem-solving.

---

[1] https://www.mturk.com
[2] https://microworkers.com
[3] https://prolific.co

Creating a Community

Individuals, researchers and policymakers can cooperate and take actions together.

Figure 15: Fostering Community Engagement among Individuals, Researchers, and Policymakers

It's crucial to create a vibrant, varied group in order to create and manage effective online lab studies. The research teams may ensure that each study reflects real-world variety and yields significant findings, by enticing a wide range of people with various backgrounds, ethnicities, and areas of expertise. This envisioned group seeks to involve users in online lab studies where they may actively participate and offer useful data. These studies can use a variety of research approaches and subject areas, making use of the online platform's ability to connect with a bigger participant pool and collect rich, varied data sets.

Stakeholders in EVIDENT seek to enable researchers to conduct their studies and collect high-quality data without being constrained by budgetary restrictions and removing the cost barrier. As a result, research projects may be more diverse, underrepresented groups may participate at higher rates, and a wider range of research subjects may be investigated.
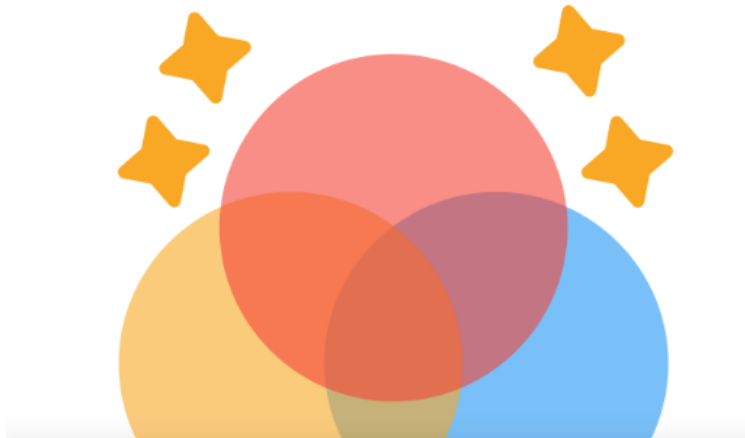
Large pool of participants

Tap into the platform's extensive participant pool, ensuring diverse and robust data for your research. Gain access to a community of engaged individuals eager to contribute to groundbreaking studies. Experience the benefits of a far-reaching audience for your experiments.

Figure 16: Accessing diverse and engaged contributors for research

Both Figure 15 and Figure 16 are on the "Home" page of the EVIDENT website [24] showcasing the importance of the aforementioned notions for the sustainability and longevity of the EVIDENT platform.

5. EVIDENT Datahub Services User Navigation

The section presents the user navigation for the datahub services as well as for additional features added in this last development cycle. The role of the datahub services is to host the available data from the project, the analytical models, the machine learning (ML) tools developed during the project's lifetime and the reports and insights derived from the results. For the platform to do so, a new module, the "Datahub Services", has been developed as described in section 3. In addition, further functionalities have been developed to support the dissemination of the collected data to Zenodo and the support of mobile applications for the supported games. Each user registered as an organisation (or organiser) can fill in their Zenodo personal access token on the user profile page[4]. After that, the user can use the platform feature to create a new deposition on the Zenodo platform automatically for each of his sessions. The rest of the section presents the user navigation of these added features.

5.1 Datahub Component

The main objective of the datahub services component is to create a place where the interested user, researchers and policymakers can find openly available data related to the EVIDENT project and the project's use cases, analytical tools and models developed by the partners through the project's lifetime and policy brief proposed based on the provided insights and during the last period of the project. To do so, any user (registered or unregistered) can visit the EVIDENT platform[5] and select the page entitled "Available data" on the main menu.

---

[4] To create a new personal access token in Zenodo, follow the steps described in the section "Creating a personal access token" on this page.
[5] https://platform.evident-h2020.eu

Figure 17: "Available data" page

A page similar to Figure 17 will present all available resources. At the time of writing, the EVIDENT platform hosts the first version of the Discrete Choice Experiment (DCE), and the Average Price Bias (APB) datasets, both datasets are related to use case 5. All open-source datasets, tools, models and reports will be eventually uploaded. For security reasons and since the datahub services are related only to the EVIDENT project, only the platform administrator is given the capability to upload new resources.

While uploading a new resource, the administrator should fill in a form with all related information, as shown in Figure 18. The form consists of seven fields described below:

- Name: The title of the new resource

- Description: The description of the new resource
- File: The file that will be available (all related formats about documents, codes and datasets are supported)
- Upload date: The date the new resource is being uploaded
- Availability: Whether the new resource will be accessible by any user or an additional justification is needed. More information about this functionality is presented in the upcoming paragraphs.
- Version: The current version of the resource[6] (e.g. Version 0.1.0)
- Type: The type of resource. We have identified three main types for our resources, tools, datasets and report; however, this categorization is easily expandable.
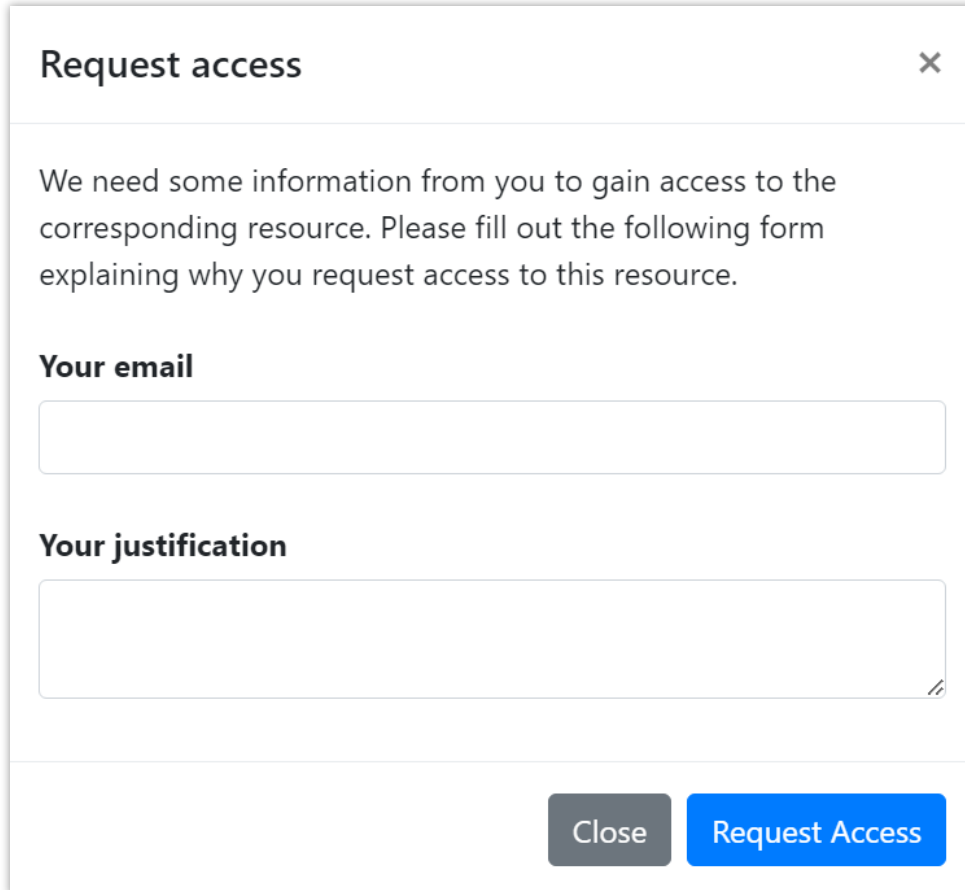
Figure 18: Add a new resource form

Resource availability

In the previous section, the concept of availability was presented. This option is used to indicate whether any resource is available or not. If the administrator selects a resource to be available, any platform user or visitor can download the corresponding resource without any restrictions. In the latter option, the platform users should provide a few pieces of information to justify gaining access to the resource. This form is presented in Figure 19.

Figure 19: Request access on resource form

If a resource is 'unavailable', a button to the previous form will be presented, similar to Figure 20.

Figure 20: Example of a resource that is "not available"

5.2   Additional features

In the previous development cycle, the EVIDENT consortium diligently worked on augmenting the functionality of the EVIDENT platform by introducing a set of additional features. This section will comprehensively elucidate the utilization and navigation of these newly incorporated features, offering valuable insights into their purpose, benefits, and seamless integration within the platform. By delving into the user navigation of these advanced features, we aim to provide users with a comprehensive understanding of their capabilities, empowering them to leverage the enhanced functionalities and optimize their experience within the EVIDENT platform.

Zenodo integration

Aiming to raise awareness about open data and encourage more users to engage with it, the EVIDENT consortium decided to leverage the Zenodo platform capabilities and integrate the provided application programming interface (API).

Zenodo provides a representational state transfer (REST) API that currently supports deposits, records and files. As a starting point, deposits refer to research outputs, records refer to published records (deposits) and files to available documents within a record. The EVIDENT platform integrates the functionalities for deposits, meaning that the organisation users are able to upload the collected replies from their sessions directly. The platform automatically drops the participants' identifiers to protect anonymity[7]. The organisation users can find this functionality in their account under the "My Sessions" page as presented in Figure 21.
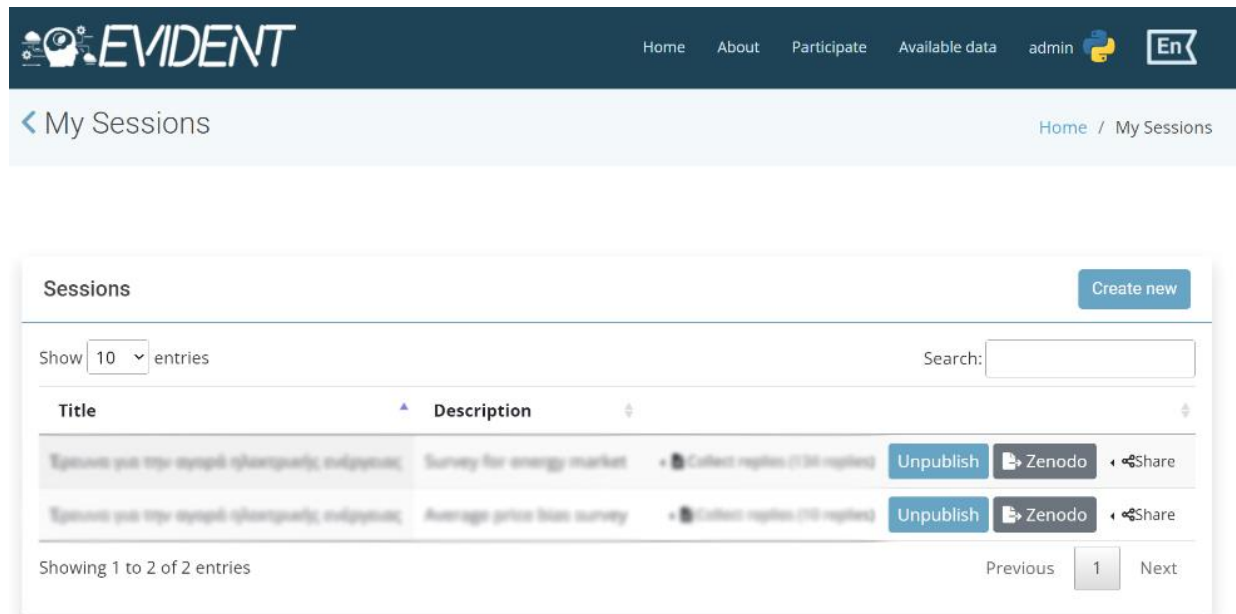

Figure 21: My sessions page - Integration with Zenodo

---

[7] A very similar functionality in terms of anonymity with the data extraction functionality described in D6.3 Gamification tools of the EVIDENT platform.

In the case where the user has not filled in their Zenodo personal access token, they will be prompted to do so. After that, the Zenodo button will open a new pop-up window requesting some actions from the user. In this step, the user can launch a new deposit, publish or delete it. In Zenodo terms, launch means that the deposit is in a draft stage and yet not published on Zenodo and publish means that the current deposit has been published on Zenodo. Finally, the user can only remove a deposit before this is published on Zenodo. This is not a limitation, but rather a feature of the Zenodo platform since every published deposit ('record') has been attributed with a unique digital object identifier (DOI).
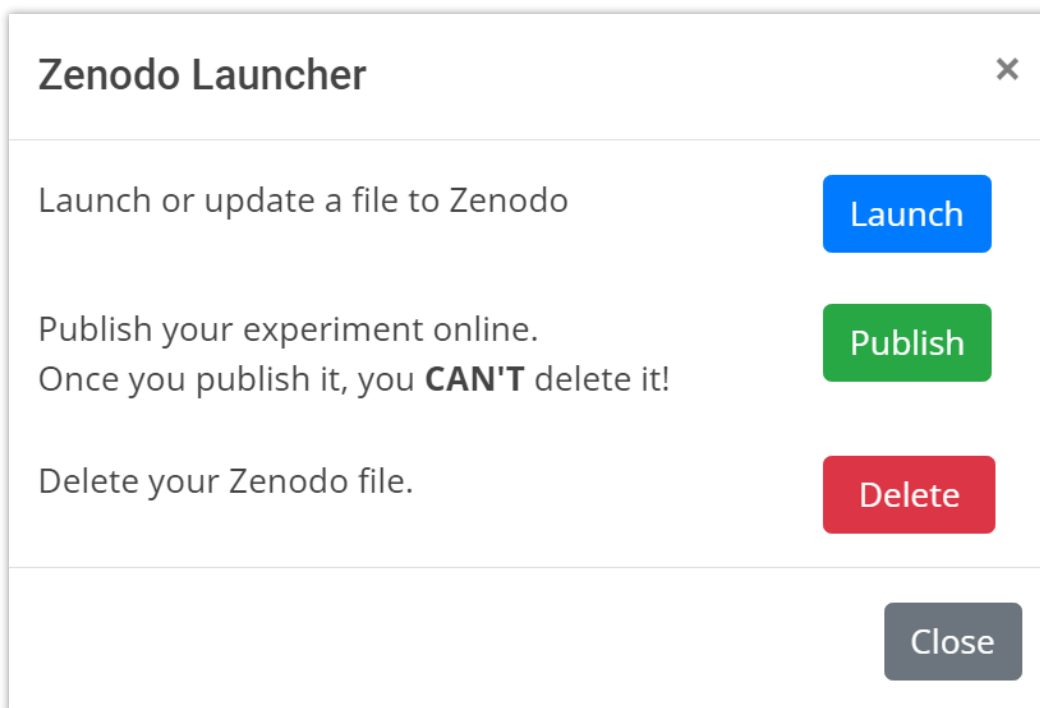

Figure 22: Share session replies to Zenodo through the EVIDENT platform

Once the user selects to launch a new deposit to Zenodo, a new form depicted in Figure 23 will appear. The introductory text on the form will notify the user that the new deposit won't get published automatically since the user himself should select to publish it (see Figure 22, Publish button). This function will create an initial deposit or update the current one. To create a new deposit, the user should fill in the following information:

- Title: the title of the deposition
- Description: the description of the deposition
- Keywords: the keywords a user can insert on the Zenodo search bar to locate the corresponding deposit
- License: the user license for the corresponding deposit

Figure 23: Launch a session to Zenodo

After creating, or updating a deposition, the user is ready to publish it on the Zenodo.

Smartphone applications and Serious Games

An additional feature added to the EVIDENT platform is the capability of a serious game to be used as a standalone smartphone application. That means the EVIDENT platform provides an API to receive session replies from 3rd party applications. For example, to maximize participation in the project's use case 4, the EVIDENT consortium developed two smartphone applications (Android and IOS) so the participants could easily join the corresponding experiment[8]. These two identical applications, a standalone version of the EVIDENT serious game, can interact with the EVIDENT platform and exchange information either as input

---

[8] The EVIDENT serious game was developed on WebGL. WebGL is a JavaScript API for rendering interactive 2D and 3D graphics within any compatible web browser. Since most mobile device web browsers are incompatible with WebGL, the participants could only join the experiment through a desktop browser.

(initialize the game for each participant) or output (collect their replies). Thus, additional fields are added in the form found in the "Add/Edit Game" pages.

Training materials for the EVIDENT platform

As a continuation of the previous section, the EVIDENT consortium aims to create additional pages within the EVIDENT platform as training materials regarding the proper elaboration of the platform on e-lab experiments. These training materials will include useful information about organisations that seek to leverage the EVIDENT platform's capabilities to design and implement e-lab experiments. In particular, the training materials will constitute a step-by-step guide about creating surveys, integrating/uploading serious games to the platform, combining them into a unified session and collecting data. The additional pages will be reported on D6.5, "Verification and Validation Report for final version of EVIDENT platform".

6. Conclusion

This section aims to highlight the key conclusions and significant achievements stemming from the development of the datahub service within the EVIDENT project. Through careful analysis and evaluation, important conclusions have been drawn regarding the functionality, effectiveness, and potential applications of the datahub service within the broader project context.

In conclusion, the datahub service of the EVIDENT platform represents a general-use and innovative solution that can host the results from the serious game applications, and surveys. Its architecture and design have been carefully developed to meet the requirements and objectives outlined in the previous work packages of the project and the specifications of the Grand Agreement (GA). The successful development and integration of the datahub component provide robust data management and promote collaboration through data-sharing functionalities.
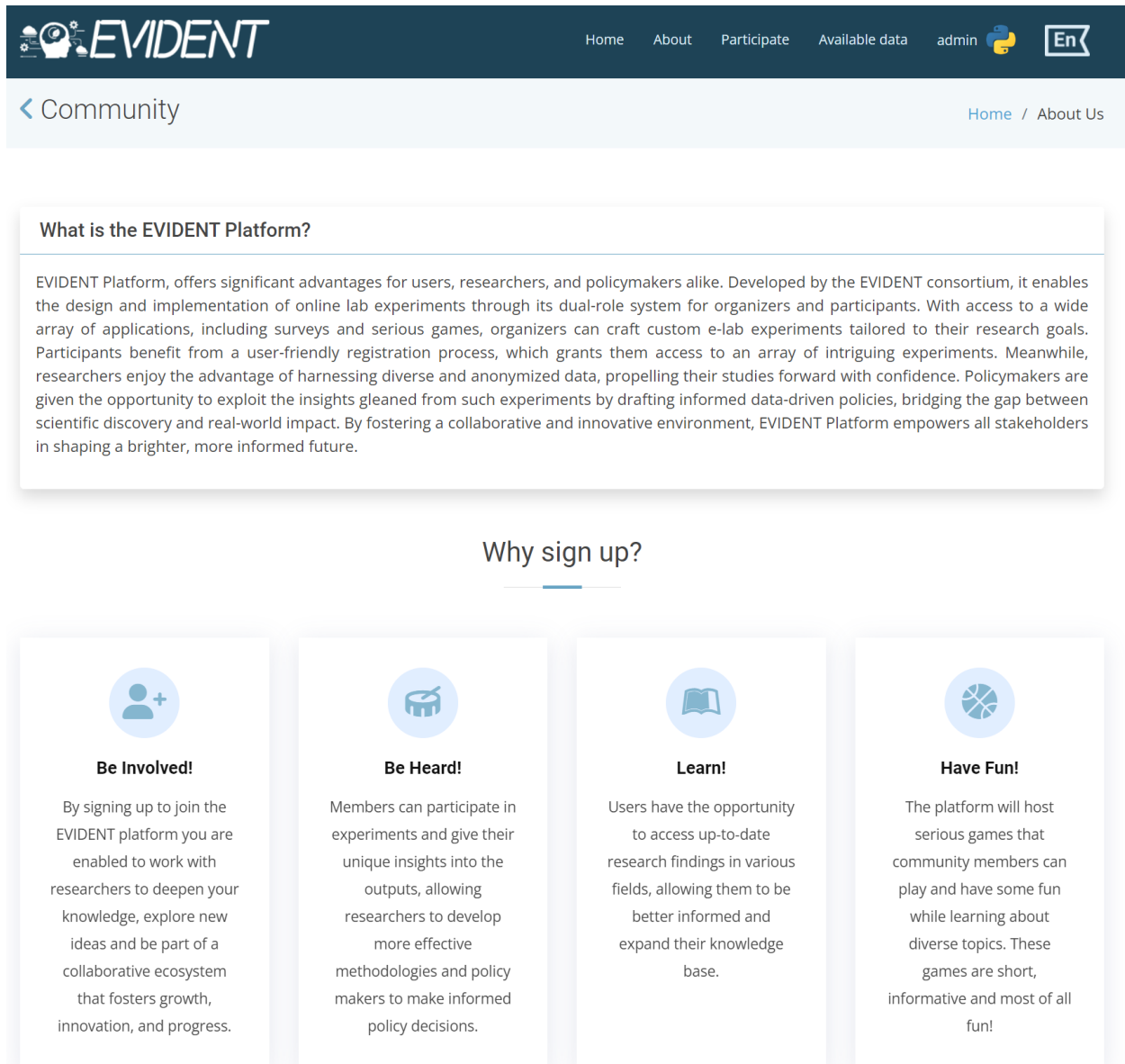
Overall, the EVIDENT platform offers some unique capabilities and advantages, making it suitable for a wide range of applications and use cases promoting collaboration, data-driven decision-making, and knowledge creation.

7. References

[1] X. Ye and P. Yue, "Financial literacy and household energy efficiency: An analysis of credit market and supply chain," *Finance Research Letters,* vol. 52, Mar. 2023.

[2] U. Batsaikhan and M. Demertzis, "Financial literacy and inclusive growth in the European Union," 2018. [Online]. Available: http://hdl.handle.net/10419/208015

[3] Amazon Web Services, "Amazon Redshift,"[Online]. Available: https://aws.amazon.com/redshift

[4] Google, "BigQuery solution," [Online]. Available: https://cloud.google.com/bigquery

[5] "Snowflake solution," [Online]. Available: https://www.snowflake.com/en/.

[6] Microsoft, "Azure Synapse Analytics," [Online]. Available: https://azure.microsoft.com/en-us/products/synapse-analytics

[7] "Teradata solution," [Online]. Available: https://www.teradata.com

[8] IBM, "Db2 Warehouse," [Online]. Available: https://www.ibm.comIBM/products/db2/warehouse

[9] Oracle, "Autonomous Data Warehouse," [Online]. Available: https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud

[10] "The Databricks Lakehouse Platform," [Online]. Available: https://www.databricks.com/product/data-lakehouse

[11] "Vertica Platform," [Online]. Available: https://www.vertica.com

[12] "Panoply Managed ELT + Cloud Data Warehouse," [Online]. Available: https://panoply.io

[13] 6Sense, "Amazon Redshift Market Share,", [Online]. Available: https://6sense.com/tech/big-data-infrastructure/amazon-redshift-market-share

[14] "Google BigQuery solution," [Online]. Available: https://cloud.google.com/bigquery

[15] 6Sense, "Market Share of Snowflake," [Online]. Available: https://6sense.com/tech/data-warehousing/snowflake-market-share

[16] 6Sense, "Market Share of Microsoft Azure Synapse," [Online]. Available: https://6sense.com/tech/big-data-analytics/microsoft-azure-synapse-market-share

[17] 6Sense, "Market Share of Teradata," [Online]. Available: https://6sense.com/tech/other-bi-and-analytics/teradata-market-share

[18] 6Sense, "Market Share of IBM Data Warehouse," [Online]. Available: https://6sense.com/tech/data-warehousing/ibm-data-warehouse-market-share

[19] 6Sense, "Market Share of Oracle Autonomous Data Warehouse," [Online]. Available: https://6sense.com/tech/data-management-and-storage/oracle-autonomous-data-warehouse-market-share

[20] 6Sense, "Market Share of Vertica," [Online]. Available: https://6sense.com/tech/big-data-infrastructure/vertica-market-share

[21] 6Sense, "Market Share of Panoply," [Online]. Available: https://6sense.com/tech/big-data/panoply-market-share

[22] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle and M. R. Munafò, "Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy," *JMIR Serious Games,* vol. 4, Jul. 2016.

[23] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Computers & Education,* vol. 59, no. 2, Sep. 2012.

[24] EVIDENT Consortium, "EVIDENT Platform Web page," [Online]. Available: https://platform.evident-h2020.eu/

Annex: Landing pages of the EVIDENT platform for individuals, researchers and policymakers.

EVIDENT

Home   About   Participate   Available data   admin   En

‹ Community   Home / About Us

What is the EVIDENT Platform?

EVIDENT Platform, offers significant advantages for users, researchers, and policymakers alike. Developed by the EVIDENT consortium, it enables the design and implementation of online lab experiments through its dual-role system for organizers and participants. With access to a wide array of applications, including surveys and serious games, organizers can craft custom e-lab experiments tailored to their research goals. Participants benefit from a user-friendly registration process, which grants them access to an array of intriguing experiments. Meanwhile, researchers enjoy the advantage of harnessing diverse and anonymized data, propelling their studies forward with confidence. Policymakers are given the opportunity to exploit the insights gleaned from such experiments by drafting informed data-driven policies, bridging the gap between scientific discovery and real-world impact. By fostering a collaborative and innovative environment, EVIDENT Platform empowers all stakeholders in shaping a brighter, more informed future.

Why sign up?

Be Involved!

By signing up to join the EVIDENT platform you are enabled to work with researchers to deepen your knowledge, explore new ideas and be part of a collaborative ecosystem that fosters growth, innovation, and progress.

Be Heard!

Members can participate in experiments and give their unique insights into the outputs, allowing researchers to develop more effective methodologies and policy makers to make informed policy decisions.
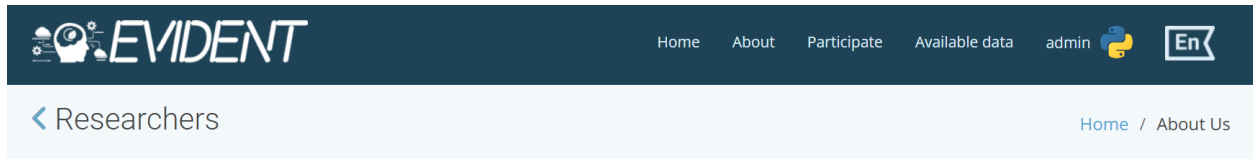
Learn!

Users have the opportunity to access up-to-date research findings in various fields, allowing them to be better informed and expand their knowledge base.

Have Fun!

The platform will host serious games that community members can play and have some fun while learning about diverse topics. These games are short, informative and most of all fun!
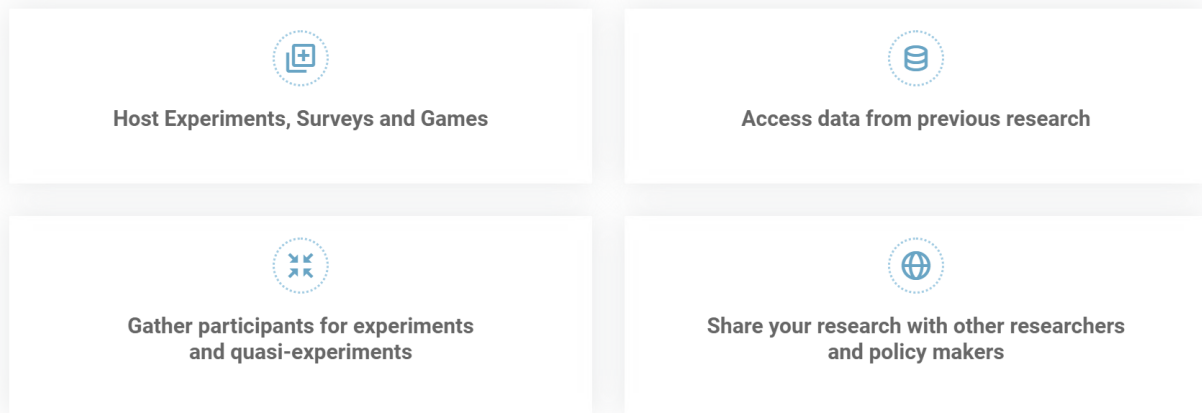
Figure 24: Landing page for individuals

Figure 25: Landing page for researchers and policymakers