Fachhochschule Nordwestschweiz
Hochschule für Technik

HackHub: Revolutionizing Data Management and Exchange in Hackatons

IP5- Project

Student	Odelia Sajtschik
Advisor	Prof. Dr. Norbert Seyff and Nitish Patkar
Client	EDHI
Project nummer	24HS_IIT23
Aggreement	confidential

Fachhochschule Nordwestschweiz, Hochschule für Technik

Windisch, April 11th 205

Summary

Energy hackathons in Switzerland serve as dynamic platforms for fostering innovation and collaboration to address critical challenges within the energy sector. These intensive events convene diverse participants, including developers, data scientists, and domain experts, to develop solutions leveraging available data. The imperative for innovation in this domain is underscored by Switzerland's ambitious energy transition goals.

The absence of a dedicated platform for managing data within these specific Swiss energy hackathons necessitates addressing key questions: What are the current access points for relevant data in the energy field? How can developers efficiently access pertinent data during these events? How can this data be effectively centralized? And how can hackathon organizers efficiently disseminate relevant data to participants?

To address these questions, the European Digital Innovation Hubs (EDIH) network organizes hackathons, and a questionnaire was administered to participants at the end of the event. The feedback obtained from this questionnaire provided essential direction on the key requirements for a data management solution. The data platform, or data hub, was subsequently developed based on this valuable input.

This research culminates in the development of a data platform designed to facilitate data sharing and reuse for hackathon challenges. The platform enables data owners to contribute, modify, and remove datasets in accordance with defined data access rules. Publicly accessible data can be readily reused across various challenges hosted on the platform. Concurrently, participants gain the capability to visualize all relevant data associated with these challenges. To enhance clarity and usability, the platform employs distinct visual markers to differentiate data types. Furthermore, datasets are directly viewable within the platform interface, contingent on their nature.

This solution could be use in the future for others field in Switzerland. Continuous development of the platform's data management functionalities may include the integration of capabilities for managing and executing algorithms directly on the hosted data.

Keywords: Energy Hackathon, Data Platform, Data Hub, Data Management, Data Sharing, Data Accessibility, Switzerland, Energy Sector.

Contents

Summary 2
Table of Figures 5
Table of Tables 5
1 Introduction 6
2. Stand of Research 8
2.1 Data Storage for the Data Hub 8
2.1.1. Comparison of Various Options for Hosting the Data 8
2.1.2 Chosen Hosting Method 14
2.2 User-Centred Requirements for Hackathon Platforms 15
2.2.1 Participant Needs and Expectations 15
2.2.2 Challenge Owner Needs and Expectations 16
3 Methods Application 18
3.1 UI and UX Design Methods 18
3.1.1 Frameworks 18
3.1.2 Requirement 19
3.1.3 Prototype 22
3.1.3 Personas 22
3.2 Data Visualisation and Management 25
3.2.1 Architecture: 25
3.2.3 Backend 29
3.2.3.1 Technologies and Frameworks 29
3.2.4 Database 32
3.3 Testing 32
3.6.1 Instrument Design 33
3.6.2 Procedure for Usability Testing – Quantitative Evaluation 33
4 Result 37
4.1 Lo-Fi Prototyping 37
4.1.1 Data Management 37
4.1.2 Data Storage with Retention Time 37

4.1.3 Form for Easy Input 38
4.1.4 Participant Data Viewing 38
4.2 Test Results.. 38
4.3 Hi-Fi Prototyping.. 39
4.3.1 Impact of Time Constraints on Prototype Evaluation 40
4.4 Summary of result.. 41
5 Conclusion.. 42
References.. 43

Table of Figures

Figure 1 Persona 1 23
Figure 2: Persona 2 24
Figure 3: Technology stack 25
Figure 4: Layers relation 26
Figure 5: Front end diagram 29
Figure 6: Backend architecture 31
Figure 7: DB tables schema 32
Figure 8: User Test part I 35
Figure 9: : User Test part II 36
Figure 10: Data sets management 37
Figure 11: New data sets form 38
Figure 12: Participants view of data sets 38
Figure 13: Owner data sets management 40
Figure 14: Participants data sets visualisation 40

Table of Tables

Table 1: On permise comparation solution 9
Table 2: Public cloud comparation solution 11
Table 3: Hybrid cloud comparation solution 13
Table 4: User frameworks 18
Table 5: Features 19
Table 6: Login functional requirement 19
Table 7: Data Visualization functional requirement 20
Table 8: Data Management functional requirement 20
Table 9: Restricted Access functional requirement 20
Table 10: System functional requirement 21
Table 11: Data Hosting functional requirement 21
Table 12: Login Non-Functional Requirement 21
Table 13: System Non-Functional Requirement 22

1 Introduction

In the field of innovation, collaborative events like hackathons play a central role as they significantly contribute to the development of novel solutions. For example, various initiatives in Switzerland foster innovation across sectors by bringing together diverse talent (Innosuisse, 2025). One key aspect of these events is the effective utilization of data, which, due to its potential to inform and drive the creation of impactful prototypes, makes a crucial contribution. The efficient management and accessibility of relevant data enable participants to generate more informed and targeted solutions (Pe-Than & Herbsleb, 2016). This allows for the development of context-aware and data-driven innovations.

Another central aspect of these collaborative events is the visualization of relevant datasets, such as energy consumption patterns, renewable energy generation, and grid load. With the growing volume of data available in the energy sector, an efficient and precise presentation of this information is becoming increasingly important to enable hackathon participants to quickly identify key insights and develop relevant solutions aligned with Switzerland's energy transition goals (Swiss Federal Office of Energy, 2022). The reports and analyses highlight the increasing importance of data-driven approaches in the energy sector (Swiss Federal Office of Energy, 2022). Through effective data management and visualization, participants can optimally leverage available information, which not only improves the quality of their prototypes but also increases the overall efficiency and impact of the hackathon.

This work focuses on the development of a generic data hub that can be used in energy-focused hackathons in Switzerland. To verify the effectiveness of this platform, a case study is examined in which various energy-related challenges are posed to hackathon participants. To ensure both accessibility and efficient data utilization, a centralized platform is developed for managing and visualizing the relevant datasets. The energy field in Switzerland is increasingly relying on advanced data analysis to optimize its infrastructure and facilitate the transition towards a more sustainable future. For example, researchers at ETH Zurich have published studies on advanced methods for maintaining the Swiss transmission grid, utilizing data-driven and deep-learning techniques for monitoring and predictive maintenance (ETH Zurich Research on Transmission Grid, 2024). Therefore, efficient data access and visualization are crucial to accelerate the development of impactful solutions during these events.

The centralized management and visualization of energy-related data make complex information understandable and intuitive for hackathon participants, especially within the limited timeframe of the event. It facilitates the rapid identification of trends and patterns, which improves the relevance and potential of the developed prototypes. Effective data management and information visualization use structured platforms and graphical representations to make large amounts of data quickly comprehensible and analysable for a diverse group of participants.

The clients, organizers of energy-focused hackathons in Switzerland, already facilitate these innovation events. However, they now want to create a centralized platform that efficiently manages, visualizes, and analyses the collected and provided energy datasets for participants across different challenges. Therefore, there is a knowledge gap and an urgent need for a new solution that addresses this challenge of fragmented data accessibility. The data hub developed in this work will serve as a central resource for future energy hackathons, streamlining data access and analysis for participants.

1 Introduction

This prototype is intended to be a user-friendly data hub where challenge owners can upload and manage datasets, and participants can easily visualize and interact with this data. This platform aims to help participants effectively understand the available energy data and develop impactful solutions within the hackathon timeframe, ultimately contributing to innovation in Switzerland's energy sector. To develop new methods, this work investigates the following research questions:

1. What are the current access points for relevant data in the energy field?
2. How can developers efficiently access pertinent data during these events?
3. How can this data be effectively centralized?
4. How can hackathon organizers efficiently disseminate relevant data to participants?

To answer the stated research questions, clear goals were formulated for this work. These goals are aimed at gaining new insights and developing practical solutions for the centralization and delivery of data for further development. The formulated goals include both quantitative and qualitative objectives that are crucial for ensuring the effectiveness and applicability of the developed systems:

Quantitative Goals:

1. To design an intuitive and user-friendly data submission process.
2. To provide a central visualization point for participants to understand challenge-specific datasets.
3. To enable the efficient reuse of public datasets across different hackathon challenges.

Qualitative Goals:

1. Participants should be able to access the dataset for a specific challenge within two clicks.
2. The error rate in data submission should not exceed 70%.
3. Conduct tests with the high-fidelity prototype with at least 5 participants.

This report provides an analysis of current options for data hubs. Initially, an overview of existing solutions for data centralization is given.

2. Stand of Research

In the development of our data hub, we focused on two critical aspects to ensure its effectiveness. First, we thoroughly researched the data itself, examining the most efficient methods for hosting and managing the data. This was essential for creating a platform that is both scalable and secure. Second, we paid close attention to the needs of the stakeholders involved in the project. Understanding their specific requirements and business goals allowed us to tailor the design and functionality of the data hub to meet their expectations. By addressing both the technical aspects of data management and the strategic needs of the stakeholders, we were able to lay the foundation for a platform that supports collaborative innovation and seamless data sharing.

2.1 Data Storage for the Data Hub

The effectiveness of our data hub, designed for integrating and harmonizing diverse datasets, fundamentally relies on a robust data storage solution. This chapter focuses on the crucial connection between the data hub and its underlying data storage architecture. It will detail the analysis undertaken to determine the optimal location for storing the data processed by the hub, considering factors such as scalability, performance, cost, and security. Ultimately, this chapter will explain the rationale behind the chosen storage strategy and how it supports the overall functionality and goals of the data hub.

2.1.1. Comparison of Various Options for Hosting the Data

The decision of where to store the data generated and managed by the data hub was a critical consideration. Several deployment models were evaluated, each with its own set of advantages and disadvantages. The primary options considered were on-premise solutions, private cloud environments, public cloud platforms, and hybrid cloud deployments. A comparative analysis of these approaches, particularly in the context of the data hub's requirements for accessibility, scalability, cost-effectiveness, and security, is presented below.

2.1.1.1 On-Premise Solutions

Description: On-premise solutions entail the acquisition, deployment, and management of data storage infrastructure within the physical boundaries of the organization's facilities (Buyya et al., 2009). This model provides the organization with direct and complete control over all aspects of the infrastructure, including hardware, software, and network configurations. While this granular control can be advantageous for specific security or compliance requirements, it also necessitates full organizational responsibility for maintenance, upgrades, security patching, and scalability (Armbrust et al., 2010).

Pros:

- **Complete Control:** Organizations have full administrative and physical control over the hardware, software, and network configurations, allowing for tailored solutions to meet specific requirements (Buyya et al., 2009).
- **Customization:** Infrastructure can be precisely configured to match specific hardware and software compatibility needs, potentially optimizing performance for workloads (Grossman, 2009).

- **Data Sovereignty:** Data remains entirely within the organization's physical infrastructure, which can be a critical factor for compliance with stringent data localization regulations and internal governance policies (Subashini & Kavitha, 2011).

Cons:

- **High Capital Expenditure:** The initial investment in hardware, licensing, and setup can be substantial, representing a significant financial commitment (Rappa, 2004).
- **Ongoing Operational Costs:** Continuous expenses for hardware maintenance, software updates, power consumption, cooling, and the salaries of specialized IT staff contribute to a significant total cost of ownership (TCO) (Knorr & Gruman, 2012).
- **Slow Deployment:** The processes of procurement, installation, configuration, and rigorous testing of on-premise infrastructure led to lengthy deployment timelines, hindering agility and responsiveness to changing needs (Hayes, 2008).
- **Scalability Challenges:** Scaling resources often requires significant lead times for hardware procurement and integration, leading to potential performance bottlenecks or over-provisioning (Armbrust et al., 2010).

Internal Comparison:

Table 1: On permise comparation solution

Feature	***Dell EMC***	***HPE (ProLiant, Synergy)***
Performance	High (e.g., Dell PowerEdge servers offer robust processing power and I/O capabilities suitable for demanding workloads [Dell Technologies, n.d.])	High (e.g., HPE ProLiant Gen10+ servers are engineered for performance and reliability with advanced memory and processing technologies [Hewlett Packard Enterprise, n.d.])
Scalability	Scaling typically involves manual procurement and integration of additional hardware, leading to significant capital expenditure and potential downtime (Hayes, 2008).	HPE Synergy offers a composable infrastructure approach, allowing for more modular scaling, but physical limitations and procurement timelines still apply (HPE, n.d.).
Management	Requires a dedicated team of skilled IT administrators for setup, configuration, monitoring, and troubleshooting (Knorr & Gruman, 2012).	Similar requirements for skilled IT personnel are necessary for managing HPE infrastructure, including firmware updates and hardware maintenance (HPE, n.d.).
Deployment Speed	The procurement, installation, configuration, and testing of on-premise hardware can span several weeks to months, depending on the complexity and scale (Grossman, 2009).	Similar deployment timelines are typical for HPE on-premise solutions, involving hardware delivery and on-site setup (HPE, n.d.).
Security	Offers a high degree of control over physical and logical security measures, as the infrastructure resides within the	HPE servers offer various security features, but the overall security posture is determined by the

	organization's secure environment (Subashini & Kavitha, 2011).	organization's implementation and policies (HPE, n.d.).
Cost	Involves substantial upfront capital expenditure (CAPEX) for hardware acquisition, followed by ongoing operational expenses (OPEX) for maintenance contracts, power consumption, and cooling (Rappa, 2004).	Similar high upfront investment and continuous maintenance costs are associated with HPE on-premise solutions (HPE, n.d.).

2.1.1.2 Private Cloud

Description: A private cloud establishes a cloud computing environment dedicated to a single organization, offering services and infrastructure that are not shared with other users (National Institute of Standards and Technology, 2011). This model can be deployed on infrastructure owned and managed by the organization (on-premise private cloud) or by a third-party provider (externally hosted private cloud), aiming to deliver the benefits of cloud computing—such as elasticity, self-service provisioning, and scalability—while providing enhanced control over resources and security (Buyya et al., 2009; Zhang et al., 2010). The goal is to create a cloud-like experience tailored to the specific needs and security requirements of the organization.

Pros:

- **Enhanced Security:** Dedicated logical isolation provides a higher level of security compared to public cloud environments, as resources are not shared with other organizations (Subashini & Kavitha, 2011).
- **Improved Integration:** Private clouds can be designed to seamlessly integrate with existing enterprise tools, compliance frameworks, and identity management systems (Hurwitz et al., 2009).
- **Flexible Networking:** Offers greater control over network configurations, allowing for the implementation of specific security policies and connectivity options, such as private links and VPNs (Vaquero & Buyya, 2011).
- **Potentially Easier Management (than on-premise):** While still requiring cloud expertise, management can be streamlined compared to the complexities of maintaining physical server infrastructure.

Cons:

- **Greater Complexity (than public cloud):** Setting up and managing a private cloud environment, especially multi-cloud solutions like Google Anthos, can be more complex than utilizing readily available public cloud services (Rimal et al., 2011).
- **Requires Specialized Expertise:** Skilled cloud professionals are necessary for the configuration, management, and security of private cloud infrastructure, including tasks like VPC peering and IAM role management (Knorr & Gruman, 2012).
- **Potential for Higher Costs:** While offering pay-as-you-go models in some cases, private clouds can incur significant costs related to network traffic, storage consumption, and backup solutions, potentially exceeding the expenses of comparable public cloud services if not carefully managed (Armbrust et al., 2010).

2.1.1.3 Public Cloud

Description: Public cloud platforms provide shared and scalable computing resources (e.g., servers, storage, networks) accessible over the internet on a pay-as-you-go basis. This model eliminates upfront infrastructure costs, with providers like Microsoft Azure, AWS, and GCP offering a wide range of on-demand services (Armbrust et al., 2010; Mell & Grance, 2011).

Pros:

- **Extremely Fast Deployment:** Resources can be provisioned within minutes using automated orchestration tools, significantly reducing setup time compared to traditional infrastructure (Armbrust et al., 2010).
- **Highly Scalable Infrastructure:** Public cloud platforms automatically adjust resource allocation through auto-scaling and load balancing, ensuring optimal performance under fluctuating workloads (Mell & Grance, 2011).
- **Integrated Security and Compliance:** Built-in security features, such as encryption and identity management, combined with extensive compliance certifications (e.g., ISO 27001, GDPR), provide strong protection for public-facing applications (Hashizume et al., 2013).
- **Pay-Per-Use Pricing Model:** Organizations are billed based on actual consumption, which optimizes costs, particularly for short-term or variable workloads, without requiring large upfront investments (Zhang et al., 2010).
- **Extensive Range of Services:** Providers offer a broad ecosystem of services for data analytics, artificial intelligence, and big data processing, enabling rapid innovation and complex data-driven initiatives (Dillon et al., 2010).
- **Global Accessibility:** Cloud services are accessible from anywhere with internet connectivity, promoting seamless collaboration and low-latency access across geographically distributed teams (Marinos & Briscoe, 2009).

Cons:

- **Limited Control Over Hardware:** Customers do not have direct access to the physical infrastructure, which can restrict customization for specific performance or compliance needs (Armbrust et al., 2010).
- **Risk of Unexpected Billing:** Without careful monitoring and management, organizations may experience cost overruns due to unanticipated resource usage or misconfigured services (Hosseini et al., 2010).
- **Need for Specialized Expertise:** Successful use of public cloud services requires cloud-specific knowledge in areas such as architecture design, security configuration, and access management, potentially increasing training or hiring costs (Rimal et al., 2010).

Internal Comparison:

Table 2: Public cloud comparation solution

Feature	***Microsoft Azure***	***Amazon Web Services (AWS)***	***Google Cloud (GCP)***
Data Pipelines	Azure Data Factory: A graphical, easy-to-use tool for building, scheduling, and	AWS Glue: A powerful, serverless ETL service. It offers flexibility but is harder	Google Cloud Dataflow: A scalable, serverless service for complex data processing. It uses Apache Beam and suits

	managing data workflows. It supports a wide range of connectors (Microsoft, n.d. [a]).	to set up and may require coding skills (Amazon Web Services, Inc., n.d. [a]).	advanced data engineering tasks (Google Cloud, n.d. [a]).
Ease of Use	High: Offers an intuitive portal, good documentation, and many templates that speed up deployments (Microsoft, n.d. [b]).	High (for experienced users): Extensive service options but requires knowledge of AWS concepts like IAM, EC2, and S3 (Amazon Web Services, Inc., n.d. [b]).	Medium (DevOps/Kubernetes focus): Developer-focused and strongly integrated with Kubernetes. Best for teams with DevOps expertise (Google Cloud, n.d. [b]).
Monitoring	Azure Monitor: Provides full monitoring across resources with metrics, logs, and alerts. Integrates well with other Azure services (Microsoft, n.d. [c]).	AWS CloudWatch: Offers unified monitoring for AWS resources and applications, including metrics, logs, and events (Amazon Web Services, Inc., n.d. [c]).	Google Cloud Operations (Stackdriver): A suite for monitoring, logging, and tracing GCP and multi-cloud environments (Google Cloud, n.d. [c]).
Pricing	Affordable: Competitive pricing with flexible options like pay-as-you-go and reserved instances. Offers detailed cost calculators (Microsoft, n.d. [d]).	Variable: Competitive pricing, but costs can rise quickly if resources are not carefully managed (Amazon Web Services, Inc., n.d. [d]).	Cost-effective for compute tasks: Known for low prices on compute-heavy workloads and discounts for sustained use (Google Cloud, n.d. [d]).
Security Compliance	Comprehensive: Meets certifications like ISO 27001, GDPR, and SOC 2. Includes services like Azure Security Center and Sentinel (Microsoft, n.d. [e]).	Extensive: Certified for standards such as FedRAMP, ISO 27001, and HIPAA. Provides security services like IAM, GuardDuty, and Security Hub (Amazon Web Services, Inc., n.d. [e]).	Robust: Complies with ISO 27001, GDPR, and CCPA. Offers tools like Cloud Armor, Security Command Center, and Cloud IAM (Google Cloud, n.d. [e]).

2.1.1.4 Public Cloud

Description: A hybrid cloud environment integrates on-premises infrastructure with public cloud resources, creating a unified system. It combines the benefits of both environments, offering greater

control over sensitive data while leveraging the scalability and flexibility of the public cloud.
Pros:

- **Control & Scalability:** Combines control of on-premises infrastructure with the scalability of the public cloud.
- **Compliance Needs:** Ideal for organizations with strict data sovereignty or compliance needs.
- **Investment Leverage:** Enables use of existing on-premises investments while extending capabilities with the cloud.

Cons:

- **Complexity:** High complexity in configuring and managing hybrid environments.
- **Cost Increase:** Increased overall cost due to maintaining both on-premises infrastructure and cloud services.
- **Deployment Time:** Potentially longer deployment times due to integrating disparate environments.
- **Latency:** Possible latency issues depending on the connectivity between on-premises and cloud resources.

Internal comparation:

Table 3: Hybrid cloud comparation solution

Feature	***Azure Arc***	***AWS Outposts***
Integration	Extends Azure services to any hardware, including on-premises infrastructure and third-party clouds, ensuring consistent management across hybrid and multicloud environments (Microsoft, 2025).	Extends AWS services to on-premises environments, integrating AWS's infrastructure into on-site data centers, offering seamless management within the AWS ecosystem (Amazon Web Services, 2025).
Ease of Deployment	Medium complexity for organizations already using Azure, as it leverages existing Azure services and tools for hybrid deployment (Microsoft, 2025).	High complexity; requires hardware delivery and installation of AWS-managed infrastructure, increasing setup time (Amazon Web Services, 2025).
Cost	Pay-as-you-go pricing model for Azure services, with additional costs for on-premises infrastructure (Microsoft, 2025).	High cost due to the capital expenditure for hardware installation, along with ongoing operational costs for AWS services (Amazon Web Services, 2025).
Flexibility	Strong flexibility, supports multicloud and hybrid configurations, with native integration for Kubernetes clusters and Azure services (Microsoft, 2025).	Strong flexibility within the AWS ecosystem, but it is limited to AWS-native services and does not support multicloud or hybrid configurations (Amazon Web Services, 2025).
Use Case	Ideal for hybrid applications, compliance-heavy organizations, and businesses with complex hybrid or multicloud needs (Microsoft, 2025).	Best for applications requiring ultra-low latency or high-performance computing, such as in edge computing or sensitive workloads (Amazon Web Services, 2025).

2.1.2 Chosen Hosting Method

This concluding section revisits the data hosting solutions previously compared, detailing the key limitations and specific inadequacies that led to their exclusion as viable options for our particular needs in the DataHub.

2.1.2.1 On-Premise Solutions

The inherent characteristics of on-premise solutions present significant drawbacks for the proposed data hub. The critical need for rapid deployment to facilitate timely access for hackathon participants and the requirement for flexible scalability to accommodate potentially volatile data volumes and user activity are fundamentally misaligned with the protracted timelines and rigid nature of on-premise infrastructure (Grossman, 2009; Hayes, 2008). Furthermore, the substantial upfront and ongoing financial commitments associated with hardware acquisition, maintenance, and dedicated IT staffing are not economically justifiable for a project demanding agility and adaptability, particularly for a potentially short-term event like a hackathon where resource efficiency is paramount (Rappa, 2004; Knorr & Gruman, 2012).

2.1.2.2 Private cloud

While private cloud solutions offer a compelling middle ground between the control of on-premise infrastructure and the flexibility of public clouds, they introduce unnecessary complexity and potential cost overhead for the specific requirements of the data hub. The primary drivers for the data hub are rapid deployment and the ability to handle potentially fluctuating demands with ease. These needs are more effectively addressed by the inherent scalability and straightforward usability of public cloud platforms. The added complexity of managing a dedicated private cloud environment, particularly a hybrid solution, would divert valuable resources from the core objectives of the data hub without providing significant advantages in terms of security or control that cannot be adequately addressed within a well-architected public cloud environment.

2.1.2.3 Public cloud

The public cloud paradigm was chosen as the optimal approach for the Data Hub initiative due to its alignment with several key requirements. These include the need for agility and elastic scalability to accommodate variable data ingestion rates and fluctuating user access patterns. The cost-efficiency inherent in the public cloud's metered billing model allows for optimal resource management, while its global accessibility ensures that the Data Hub can operate seamlessly across different geographic locations. Furthermore, the robust ecosystems provided by cloud platforms such as Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform (GCP) offer advanced services in data storage, ETL processes, real-time analytics, and machine learning. These capabilities not only meet the current needs of the Data Hub but also provide a foundation for future scalability and expansion.

Among the public cloud options, Microsoft Azure was selected primarily due to the team's existing familiarity and expertise with the platform. This prior knowledge of Azure's tools and services facilitated a smoother implementation process, ensuring that the Data Hub could be developed efficiently and effectively.

2.1.2.4

A hybrid cloud's complexity and cost are unnecessary for the data hub. Its core needs—global accessibility and scalable storage—are efficiently fulfilled by a well-architected public cloud (Mell, P., & Grance, T., 2011). Public clouds offer geographically distributed, scalable, and highly available infrastructure ideal for worldwide data access. The added management overhead, network complexities, and costs of a hybrid setup offer no significant advantage for this use case, which aligns perfectly with the streamlined and cost-effective nature of a pure public cloud. An on-premises component would likely increase latency and management burdens without addressing the data hub's fundamental requirements.

2.2 User-Centred Requirements for Hackathon Platforms

In developing the DataHub platform, it is crucial to ensure that the needs of its key stakeholders—*challenge owners* and *participants*—are met. These two user groups have distinct roles and expectations within a hackathon, and understanding these differences is vital for creating an effective and engaging platform. A thorough analysis of participant expectations, coupled with a review of the business goals outlined by the challenge owners, informs the platform's design to support these users effectively (). The following sections outline the specific needs and expectations of both groups.

2.2.1 Participant Needs and Expectations

Participants are individuals or teams who engage in the hackathon by selecting and solving challenges set by the challenge owners. Their role is to develop creative, innovative solutions within the constraints of the hackathon environment. For participants, the platform should provide a seamless experience that enables them to focus on problem-solving, collaboration, and development. The primary user stories highlighting their needs are as follows:

- *Seamless Registration and Onboarding*
  Participants expect a quick and simple registration process. Social login options and clear guidance about the platform’s purpose and upcoming events are essential for easy onboarding. After registration, participants should be directed to a user-friendly dashboard with a checklist or timeline, helping them stay organized and informed about key milestones during the hackathon.
- *Efficient Challenge Discovery*
  Upon logging in, participants need an easy way to browse available challenges. The platform should present clear previews of challenges, showcasing the challenge titles, themes, and brief descriptions. Participants should have access to detailed challenge pages, which include pitch slides, requirements, and other resources that help them make informed decisions about which challenge to join.
- *Comprehensive Challenge Information*
  Once a challenge is selected, participants require in-depth information to guide their efforts. This includes a complete description of the challenge, details about the technology stack and tools needed, team role requirements, and clear instructions on how to access the datasets. A standardized ReadMe file with preparation guidelines, data access instructions, and onboarding materials is essential. Additionally, participants expect open communication channels (e.g., Slack groups) and welcome messages from the challenge owners to facilitate smoother integration into their teams.
- *Streamlined Data Access and Exploration*
  Data is a core component of hackathons, and participants need straightforward access to the

datasets required for their challenges. The platform should provide centralized access to datasets, with the ability to preview and explore them. A robust search and filtering system should help participants find datasets based on specific tags, relevance, and data type, ensuring they can work efficiently and make the most of available resources.

- *Integrated Collaboration Tools*
  Collaboration is essential in hackathons. The platform should offer integrated tools, such as Kanban boards and task lists, to facilitate team coordination. Built-in communication tools, like chat channels, are necessary for seamless interaction between team members. A live event dashboard to track announcements and deadlines will ensure participants stay informed and organized throughout the event.
- *Clear Submission Process*
  At the end of the hackathon, participants need a simple and clear submission process. The platform should offer a submission interface that allows for previewing, confirming, and submitting their solutions with ease. Real-time acknowledgment or feedback on submissions reassures participants that their work has been received and is under review.
- *Post-Hackathon Engagement*
  After the hackathon concludes, participants value continued access to their workspaces, archived datasets, and opportunities for networking. The platform should offer a community space for participants to engage with others, share progress, and build on their ideas. Post-event summary emails with useful resources and next steps are also appreciated to help participants remain engaged with the event's outcomes.

2.2.2 Challenge Owner Needs and Expectations

Challenge owners are the individuals or organizations responsible for creating, managing, and overseeing the challenges in the hackathon. Their role is to define the problems that participants will solve, provide the necessary resources (such as datasets), and ensure the overall smooth operation of the event. The following user stories highlight the key needs and expectations of challenge owners:

- *Easy Challenge Creation and Management*
  Challenge owners need a platform that allows them to easily create and manage challenges. The platform should feature a guided challenge creation process, where owners can define the challenge, set objectives, and upload necessary resources like datasets or APIs. A user-friendly dashboard is essential for efficiently managing challenge descriptions, team role requirements, and specific challenge criteria.
- *Secure Data Provision and Control*
  Data security is a key concern for challenge owners. The platform must provide secure data upload and management features, allowing challenge owners to control who has access to the datasets. Granular control over data privacy (e.g., public versus private settings) ensures that sensitive data is handled appropriately. Challenge owners also benefit from the ability to categorize challenges, which helps participants discover relevant challenges easily.
- *Effective Participant Onboarding*
  To ensure participants are well-prepared, challenge owners need the ability to upload and share onboarding materials, such as ReadMe files, data access instructions, and event preparation guides. Providing clear onboarding resources ensures that participants understand the expectations and have the necessary information to begin working on the challenge.
- *Seamless Collaboration and Communication*
  Communication between challenge owners and participants is crucial throughout the

hackathon. The platform should facilitate communication via built-in messaging systems, allowing challenge owners to directly interact with teams, answer questions, and provide feedback. Announcements and updates can be shared easily through the platform, ensuring that all participants are kept informed about important changes or deadlines.

- *Efficient Data and Process Management*
  During the event, challenge owners may need to upload additional resources, datasets, or updates. The platform should provide tools for challenge owners to easily manage these updates and ensure participants have access to the latest information. Additionally, challenge owners need a system for reviewing participant submissions, providing feedback, and monitoring the progress of teams to ensure the event runs smoothly.
- *Post-Hackathon Engagement*
  After the hackathon, challenge owners benefit from continued access to the platform for gathering feedback and refining solutions. They should be able to showcase winning projects, share post-event resources, and solicit feedback from participants. Continued engagement ensures that the hackathon has a lasting impact and can lead to further collaboration or opportunities for development.

3 Methods Application

This chapter describes the methods and data used in this work. It begins with an overview of the UI and UX design methods, followed by visualization methods and the technology used. Subsequently, the methods for determining the ranges and data aggregation are discussed, as well as the eye tracking and usability testing.

3.1 UI and UX Design Methods

This chapter examines the field of UI/UX and the impact of design concepts. Personas were created to better understand the usage context. Wireframes were also generated in the form of Lo-Fidelity (Lo-Fi) and High-Fidelity (Hi-Fi) prototypes, with the aim of creating a product that is as realistic as possible and can be implemented.

3.1.1 Frameworks

The goal is to create a positive user experience through the use of visualizations and to evoke positive emotions in the users. To achieve this goal, some of the most used design frameworks were examined. Table 2 also shows which task areas of the project work can be assigned to the frameworks:

Table 4: User frameworks

Framework	***Description***	***Use***
Design Thinking	An iterative process that utilizes creativity and critical thinking to solve complex problems (Dam & Siang, 2024).	Ensures that solutions are both technologically and economically feasible and meet user needs.
		Initially, hand sketches and wireframe tools like Figma were used, and these were subsequently implemented in Java and Javascript.
User-Centered Design (UCD)	Emphasizes the importance of the user perspective and its integration into every phase of the design process (Ferris, 2004).	Creates solutions that are intuitive and easy to use through constant user involvement and feedback.
		Personas, Usability Testing, and Data visualisation.
Kanban	as a lean, agile framework for managing workflow through visualization and limitation of work in progress. (Anderson, 2010)	Managing the work by visualizing flow, limiting work in progress, and continuously improving the process.
		Backlog of work items visualized on the board to manage their flow through the system.

To ensure a user-friendly outcome focused on positive emotional engagement through visualizations, Kanban framework was chosen for this project. This approach allows for the visualization of user-centric design tasks, such as incorporating usability feedback and refining interfaces based on user data. The iterative nature of Kanban, with its emphasis on continuous improvement, will facilitate a development process that prioritizes ease of use and aims to evoke positive user emotions.

3.1.2 Requirement

Customer needs are the focus of this chapter. The Customer Hierarchy of Needs method (Soares et al., 2022) helps identify them. This method has four categories of customer needs: Functional Needs (basic purpose fulfillment), Operational Needs (smooth usage/management), Value Needs (additional customer benefit), and Efficiency Needs (performance/resource optimization).

3.1.2.1 Features:

Table 5: Features

ID	Name	Description
F01	Login	Allows registered users (owners and participants) to securely log in to the DataHub using their credentials, with role-based redirection and access control.
F02	Data Visualization	Provides the ability for participants to view datasets in a clear and understandable tabular format.
F03	Data Management	Enables owners to add, edit, and delete datasets associated with challenges they manage, including the integration of public datasets.
F04	Restricted Access	Ensures that users can only access the data and functionalities relevant to their assigned role (participant or owner) and their specific challenge associations (joined or owned).
F05	System Reaction	Defines how the system responds to user actions (e.g., redirects, updates, feedback), tailored to the user's role and the context of the interaction.
F06	Data Hosting	Describes the system's capability to securely store and serve the datasets and related information, with access control based on user roles and permissions, and provisions for backup and scalability.

Functional Requirement

Table 6: Login functional requirement

ID	Description
F01.R01	The system shall allow registered users to enter their username and password.
F01.R02	The system shall authenticate the user's credentials against the user database and identify their role (owner or participant).
F01.R03	Upon successful authentication, the system shall establish a secure session for the user, specific to their identified role.
F01.R04	If the user's role is "owner," the system shall redirect them to the list of challenges they manage.

F01.R05	If the user's role is "participant," the system shall check if they have joined any challenges. If not, redirect to the available challenges list; otherwise, redirect to the dataset of their joined challenge.
F01.R06	Upon failed authentication, the system shall display an appropriate error message to the user.

Table 7: Data Visualization functional requirement

ID	Description
F02.R01	The system shall adapt the representation of the dataset based on its format: display CSVs as tables, PDFs in a PDF viewer, and web apps in an embedded frame.
F02.R02	The system shall provide pagination controls for participants to navigate through large datasets.
F02.R03	The system shall allow participants to sort the data in the table by clicking on column headers.
F02.R04	Participants shall have view-only access; editing or deleting datasets is restricted to owners.

Table 8: Data Management functional requirement

ID	Description
F03.R01	Owners shall be able to add new datasets to a challenge they manage, providing necessary information (title, description, organization, format, link).
F03.R02	Owners shall be able to edit the details of existing datasets associated with challenges they manage.
F03.R03	Owners shall be able to delete datasets from the challenges they manage.
F03.R04	When adding a new dataset to a challenge they manage, owners shall be able to view and select from a list of "public" datasets.
F03.R05	Owners shall be able to add selected "public" datasets to challenges they manage without re-entering their information.

Table 9: Restricted Access functional requirement

ID	Description
F04.R01	The system shall ensure that participants can only access the datasets of the challenges they have joined.
F04.R02	The system shall ensure that owners can only manage the challenges they have created and the datasets associated with those challenges.

F04.R03	The system shall prevent participants from accessing data management functionalities (add, edit, delete datasets).
F04.R04	The system shall prevent owners from accessing or modifying challenges or datasets that they do not own.

Table 10: System functional requirement

ID	Description
F05.R01	Upon successful login, the system shall provide visual feedback to the user confirming the login and indicating their role.
F05.R02	After a participant joins a challenge, the system shall provide a confirmation message and redirect them to the associated dataset view.
F05.R03	When an owner adds, edits, or deletes a dataset, the system shall provide a confirmation message indicating the success or failure of the action.
F05.R04	The system shall display appropriate error messages to users when actions cannot be completed, tailored to their role and the context.
F05.R05	The system shall maintain a consistent user interface and provide clear feedback on user interactions, consistent with the user's role.

Table 11: Data Hosting functional requirement

ID	Description
F06.R01	The system shall provide secure storage for all datasets and related metadata, with access controls based on user roles and challenge associations.
F06.R02	The system shall ensure the availability of the stored data to authorized users based on their role and permissions.
F06.R03	The system shall be capable of handling datasets.

Non- Functional Requirement

Table 12: Login Non-Functional Requirement

ID	Description
F01.R01	User credentials must be stored using encrypted passwords (e.g., bcrypt).
F01.R02	The login functionality must be available 24/7 with a target availability of 99.9%.
F01.R05	The codebase must follow clean architecture principles to ensure maintainability and scala

Table 13: System Non-Functional Requirement

ID	Description
F05.R06	The codebase should follow clean architecture principles to ensure maintainability, scalability, and testability.

3.1.3 Prototype

3.1.3.1 Low-Fidelity

Low-fidelity wireframes are basic visual layouts of a user interface, intentionally lacking intricate design specifics. They bridge the gap between user research findings and the eventual visual design by outlining a product's fundamental structure, content organization, and user navigation pathways (Green, 2024). For this project, initial concepts were developed using Figma. This approach enabled rapid and straightforward presentation of ideas to clients, facilitating the incorporation of their feedback. Constructing these Lo-Fi wireframes was a crucial preliminary stage for developing the more detailed Hi-Fi wireframes.

3.1.3.2 High-Fidelity

High-fidelity prototypes are intricate product mockups designed to closely mirror the final version and provide an authentic user experience. They offer stakeholders a preview, enabling them to give specific and in-depth feedback. The prior creation of Lo-Fi prototypes is a necessary step. Following their development, the intention is to gather detailed feedback to refine and finalize the design. Subsequently, the decision was made to create the Hi-Fi prototype.

3.1.3 Personas

Detailed profiles of two fictional users representing the typical characteristics and needs of the target group (adapted from Cooper et al., 2007, pp. 77–83) to focus our prototype development on certain specifics groups.

Identifying the Target Audience
Identifying the target audiences is crucial for the success of a solution. As these target audiences include various stakeholders, from developers and designers to domain experts and potential users of the hackathon outcomes, we want to specifically focus on the following two main groups, as they are central to the hackathon process:

1. Participants: Individuals (developers, designers, etc.) who will be actively working on the challenges and developing solutions.
2. Challenge Owners: The individuals or organizations who define the problems and set the goals for the hackathon challenges.

Personas enable a better understanding of the users' needs and should help in developing relevant and engaging challenges and solutions. The following is a profile of a hackathon participant:

PERSONA: Dr. Anya Müller

NAME

Dr. Anya Müller

Demographic

Female 40 years

Switzerland

Data Scientist and Energy Sector Expert

Goals

Seamlessly Post and Manage Challenges: Anya wants an intuitive platform where she can easily define, update, and manage her hackathon challenges without needing constant external assistance.

Control Data Access for Privacy: Maintaining the confidentiality of sensitive energy sector data is paramount. Anya needs fine-grained control over who can access which datasets and for what purpose.

Engage with Participants Effectively: Anya wants to interact with participating teams during and after the hackathon to provide guidance, answer questions, and potentially explore further collaboration.

Quote

I need a platform where I can manage challenges and data effortlessly, without having to depend on organizers for every update.

Behavior & Attitudes:

Values Data Confidentiality: Anya prioritizes the security and privacy of data above all else. She expects any platform she uses to have robust security measures and granular control over data access.

Expects Reliable Tools: As a data scientist, Anya relies on tools that are stable, accurate, and function as expected. She has little patience for buggy software or unreliable features.

Open to Collaboration (Structured): Anya is willing to engage with hackathon participants and data partners, but she prefers a well-defined process with clear communication channels and expectations. Spontaneous or unstructured interactions are less efficient for her.

Experienced with Documentation: Anya understands the importance of clear and concise documentation and consistently creates README files and other documentation for her projects. She appreciates platforms that facilitate this.

Experienced with Basic Web Platforms: Anya is comfortable navigating and using standard web-based platforms for data management and collaboration. However, overly complex or unintuitive interfaces can be frustrating.

Communication Requires Effort (Potentially Personal Info): While willing to communicate, Anya finds it can sometimes be inefficient. The need to exchange personal information like WhatsApp numbers for effective communication is a pain point, suggesting a preference for integrated platform communication tools.

Frustrations

Reliance on Organizers: Frustration stems from the lack of direct control over her challenges and data, requiring her to rely on hackathon organizers for updates and information dissemination.

Inefficient Data Sharing: The need to sometimes use physical media like flash drives to share data highlights a lack of secure and user-friendly data sharing mechanisms.

Difficult Onboarding: The lack of standardized templates or tools for data presentation and onboarding creates extra work and slows down the initial stages of the hackathon for participants.

Motivations

Drive Innovation: Drive energy sector innovation through data-driven challenges.

Control maintenance: Maintain strict control and security over sensitive energy data.

Efficient Management: Achieve seamless and efficient management of challenges and data without relying on intermediaries.

Engagement with Participants : Foster meaningful and structured engagement with hackathon participants to guide problem-solving and gain insights.

Simpler administration: Minimize administrative burden and focus on the core substance of the challenges.

Figure 1 Persona 1

PERSONA: Elias Keller

NAME

Elias Keller

Demographic

♂ Male 28 years

Switzerland

Fullstack Developer

Goals

Easily View Challenges and Register: Elias wants a straightforward and intuitive process to browse available challenges and sign up for the ones that interest him.

Easy Access to Information: During the hackathon, Elias needs quick and reliable access to all necessary information, including challenge details, data descriptions, documentation, and rules.

Collaborate Effectively with Team Members: Elias aims for seamless collaboration with his team, requiring effective communication channels and tools for sharing ideas and progress.

Use Relevant, High-Quality Data: Elias expects the provided data to be pertinent to the challenges and of sufficient quality to enable meaningful problem-solving.

Easy Access and Re-access to Data: Once he has access to the data for a challenge, Elias needs to be able to easily retrieve it whenever needed without unnecessary hurdles.

Quote

I just want a platform where I can focus on solving challenges, not figuring out how to access tools or data.

Behavior & Attitudes:

Tech-Savvy: Elias is comfortable using various online platforms and tools, including GitHub for version control, Slack for communication, and Kaggle for data science competitions.

Prefers Integrated Communication: While familiar with external communication tools, Elias appreciates platforms that offer built-in communication features for seamless team collaboration within the hackathon environment.

Proactive and Resourceful: Elias actively seeks out resources, documentation, and tools that can help him be more productive and effectively tackle the challenges.

Values Data Accessibility: Elias wants quick and easy access to high-quality, relevant data for the challenges. He appreciates data hubs that offer partial previews to understand the data before fully committing to a download.

Frustrations

Cumbersome Registration: A lengthy or unclear registration process can be frustrating and a barrier to entry.

Lack of Integrated Tools: The absence of built-in tools for team formation, role management, and communication forces teams to rely on external platforms, which can be less efficient.

Difficult Data Access: Challenges in finding, understanding, or accessing the required data quickly can significantly hinder progress.

No Data Previews: Not being able to preview data before downloading makes it harder to assess its relevance and can waste time and resources.

Limited Post-Hackathon Support: The lack of opportunities or tools for continued collaboration or development after the hackathon ends can be disappointing.

Motivations

Participate in an efficient hackathon that minimizes logistical hurdles and maximizes focus on problem-solving.

Easily discover and register for interesting and relevant challenges.

Gain straightforward and quick access to all necessary information and high-quality data required for the challenges.

Collaborate seamlessly and effectively with team members through integrated communication and collaboration tools.

Utilize relevant and high-quality data to develop meaningful solutions and enhance his skills

.Have easy and repeated access to the challenge data without unnecessary complications.

Potentially benefit from continued collaboration or development opportunities beyond the hackathon event.

Figure 2: Persona 2

3.2 Data Visualisation and Management

This section outlines the technology stack underpinning the datahub application. Subsequently, a detailed exposition of each architectural layer will be provided, elucidating the specific technologies and components utilized within the application.

3.2.1 Architecture:

The web application for the data visualization is structured around a three-tier architecture. three distinct layers are: the Front End (Presentation Layer), the Back End (Application Logic Layer), and the Database (Data Persistence Layer). Figure 3 provides a high-level overview of this architecture and the interaction between its components.

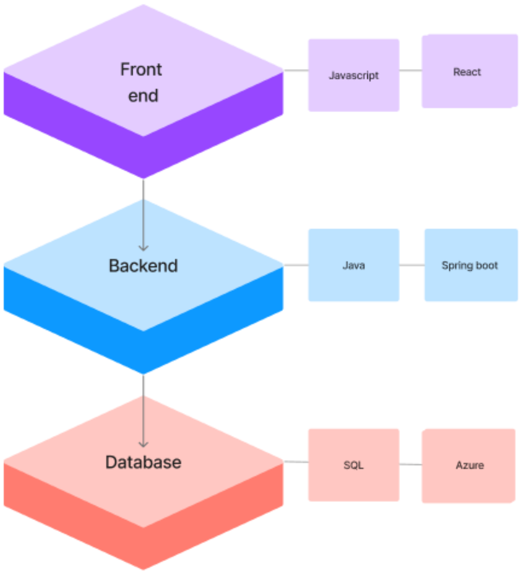


Figure 3: Technology stack

This next Figure will visualize the representations of the relation between the presentation layer, the business logic layer, and the data access layer.

3 Methods Application

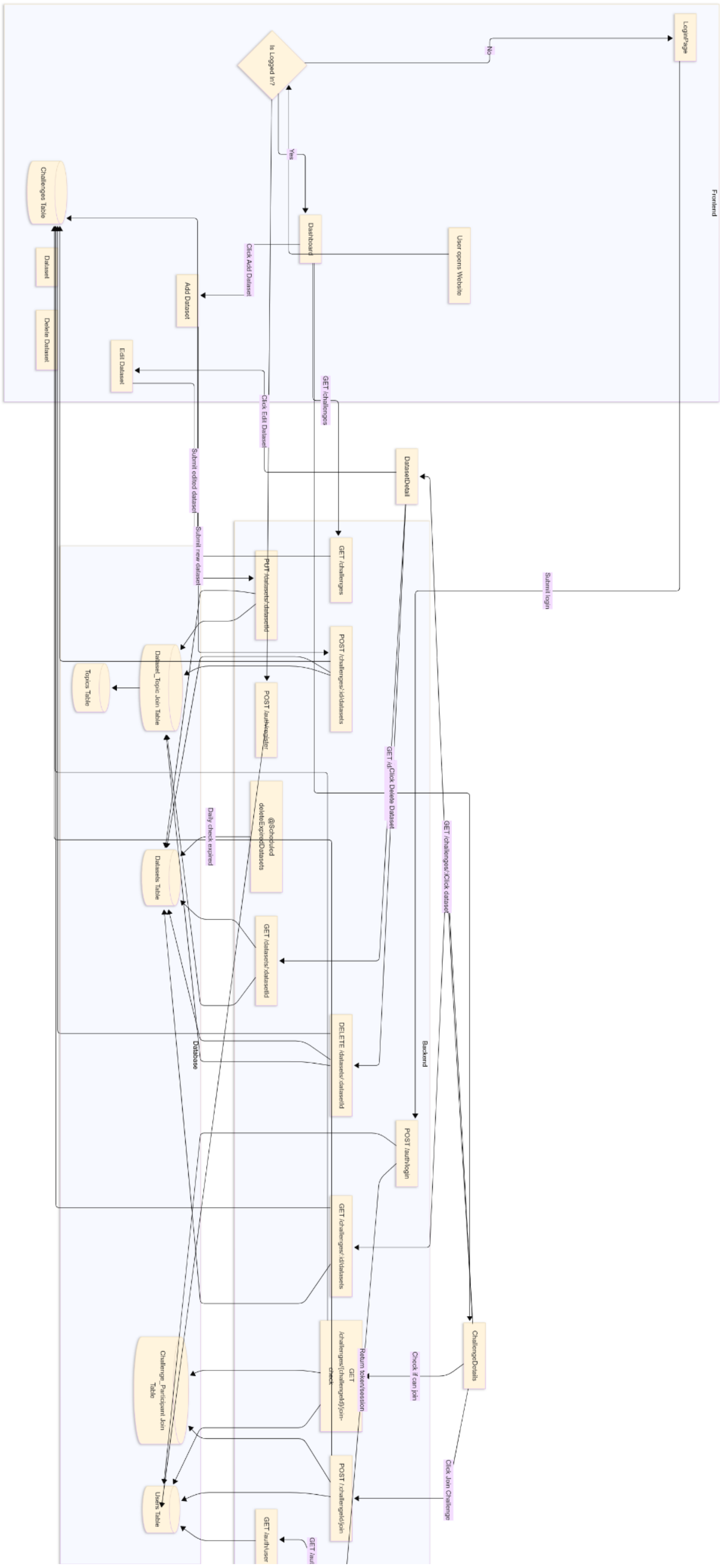


Figure 4: Layers relation

saveSession

3.2.2 Frontend:

The front end serves as the user interface, responsible for rendering interactive visualizations and handling user input.

3.2.2.1 Technologies and Frameworks

Programming Languages: JavaScript

The primary programming for the development is **JavaScript**. JavaScript is one of the most widely utilized languages in web development due to its versatility and extensive support across modern web browsers. It facilitates the creation of dynamic and interactive user interfaces, making it a fundamental technology for frontend development (MDN Web Docs, n.d.).

Advantages of JavaScript:

- **Interactivity and Dynamism:** JavaScript enables websites to respond to user input, thereby supporting interactive elements such as forms, buttons, and dynamic content updates (MDN Web Docs, n.d.).
- **Browser Compatibility:** JavaScript operates seamlessly across nearly all contemporary web browsers, enhancing the accessibility and user experience of web applications (Eich, 1995).
- **Extensive Ecosystem:** JavaScript benefits from a rich ecosystem of libraries and frameworks such as React, Angular, and Vue.js, which streamline the development process and enhance code maintainability (MDN Web Docs, n.d.).

UI Framework: React

The application's user interface was developed using React, a JavaScript library created by Facebook. React is particularly suited for building interactive user interfaces, leveraging a component-based architecture where UI elements are treated as modular, reusable components (Facebook, n.d.).

Advantages of React:

- **Component-Based Architecture:** React's component-based structure encourages modularization, allowing developers to break down complex UIs into smaller, reusable parts. This approach enhances code maintainability and facilitates easier collaboration among developers (Facebook, n.d.).
- **Virtual DOM:** React uses a virtual DOM, which optimizes updates to the user interface by minimizing direct manipulation of the actual DOM. This results in improved application performance (Facebook, n.d.).
- **Flexibility and Integration:** React can easily integrate with other libraries and tools, making it a flexible choice for building complex web applications (Abramov & Clark, 2015).

3.2.2.1 Implementation in Front end

1. **Security Aspect**
   A critical component of the application is the management of privacy for sensitive data. Following meetings, it was decided that private data would be accessible exclusively within the designated challenge group.
   Furthermore a "Store until" filed was added to the management form. Additionally,

participants are required to explicitly validate their consent before being permitted to view protected data, ensuring compliance with privacy and security guidelines.

2. **Data Reuse (Owner Page)**
   On the owner page, users will have the ability to reuse existing data.
   When entering text into the title field, the system will suggest existing titles to facilitate data reuse and improve efficiency.
3. **Data Management (Owner Page)**
   The owner page will provide specific functionalities for managing data entries, including the ability to:
   add new data, edit existing data, and delete data entries.
4. **Searching**
   A search functionality has been implemented to enable users to locate specific data quickly and efficiently.
   The search system is designed to provide relevant results in response to user queries.
5. **Data Visualization**
   Data will be sourced from multiple forms and provided via links.
   After further discussions, it was decided that the application would visualize data or images by rendering them directly from the provided links, ensuring a streamlined and user-friendly display of content.

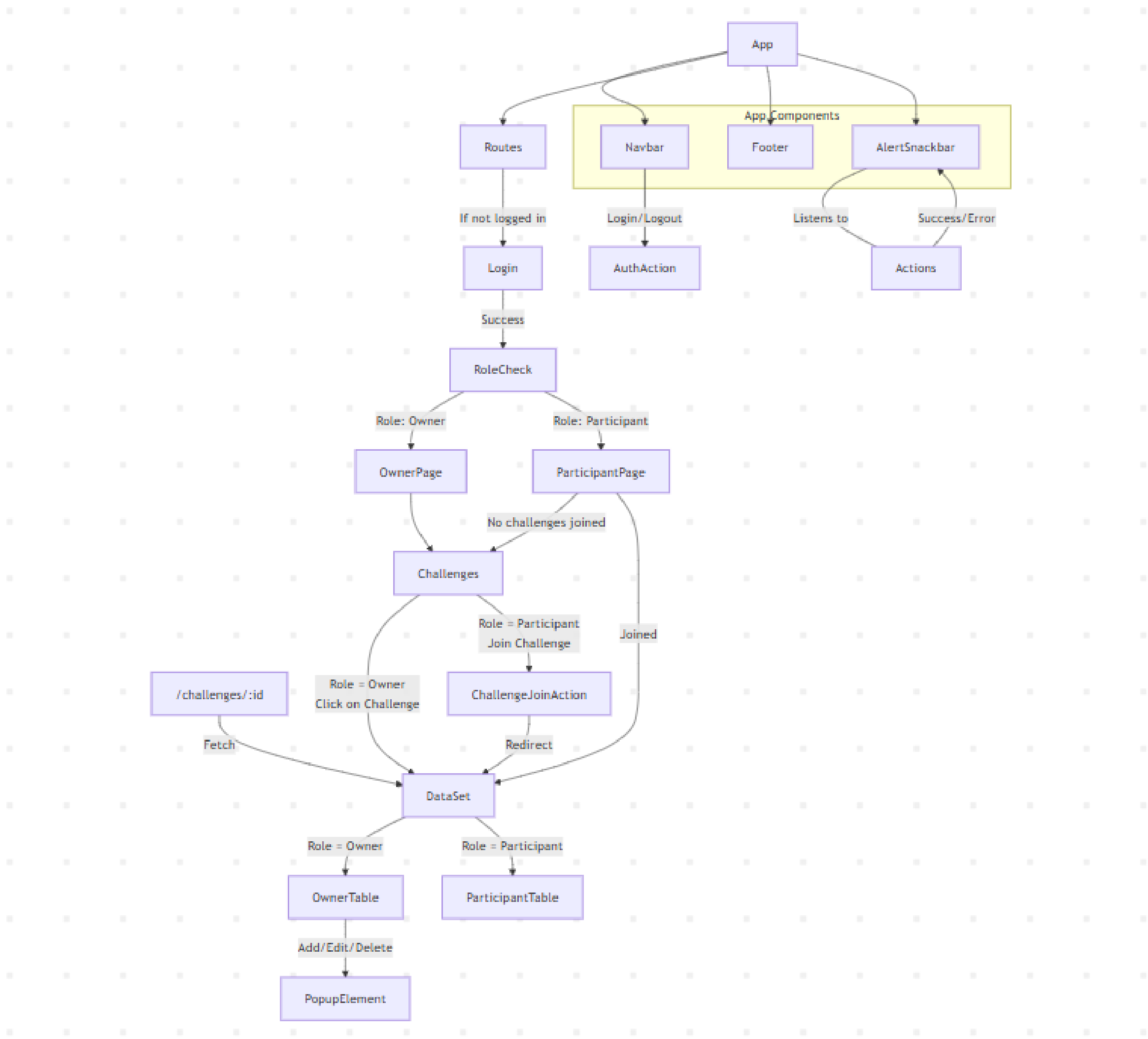


Figure 5: Front end diagram

3.2.3 Backend

The back end of the data visualization application is responsible for handling data, managing application logic, and providing an API for the front end to consume. Built using Java and the Spring Boot framework, this layer ensures the secure and efficient operation of the application.

3.2.3.1 Technologies and Frameworks

Programming Language: Java

The primary programming language utilized for the back-end development is Java. Java is a robust, platform-independent, and widely adopted language in enterprise application development. Its strong typing and extensive libraries make it well-suited for building scalable and maintainable server-side applications (Oracle, n.d.). Advantages of Java:

- **Platform Independence:** Java's "write once, run anywhere" capability allows the back end to be deployed on various operating systems without code modification (Gosling et al., 1996).

- **Scalability and Performance:** The Java Virtual Machine (JVM) is highly optimized for performance, and the language provides features for building concurrent and scalable applications (Bloch, 2018).
- **Rich Ecosystem:** Java boasts a vast ecosystem of libraries and frameworks, including Spring Boot, which significantly simplifies and accelerates the development process (Walls, 2019).

Framework: Spring Boot The application's back end is built using Spring Boot, a powerful and widely adopted Java framework. Spring Boot simplifies the process of building stand-alone, production-grade Spring-based applications with minimal configuration. It provides numerous features out-of-the-box, such as dependency injection, auto-configuration, and embedded server support (Spring.io, n.d.). Advantages of Spring Boot:

- **Simplified Configuration:** Spring Boot auto-configures much of the application infrastructure, reducing the need for verbose XML configurations and accelerating development (Walls, 2019).
- **Rapid Development:** Its convention-over-configuration approach and various starter dependencies streamline the development process, allowing developers to focus on business logic (Spring.io, n.d.).
- **Robust Ecosystem Integration:** Spring Boot seamlessly integrates with other Spring projects and a wide range of third-party libraries,providing flexibility and power for building complex back-end systems (Spring.io, n.d.).
-

3.2.3.1 Implementation in Back end

1. **Security and Authentication :**Spring Security handles authentication. The SecurityConfig class configures it, and CustomAuthenticationManager validates credentials and issues tokens (SecurityConfig and CustomAuthenticationManager).
2. **User Roles:** The User class defines Owner and Participant roles. Access control is enforced using @PreAuthorize in controllers (User class and SecurityConfig).
3. **Dataset Management:** The DatasetService class allows Owners to add, edit, or delete datasets with methods like addDataset(), updateDataset(), and deleteDataset() (DatasetService and DatasetController).
4. **Challenge Participation:** Participants join challenges, and dataset access is based on challenge membership. This is managed by joinChallenge() and checkUserJoinedChallenge() (ChallengeService and ChallengeController).
5. **Data Integrity and Protection:** Data consistency is ensured using transactions in service methods. Inputs are validated in DTOs like DatasetDTO and ChallengeDTO (DatasetDTO and ChallengeDTO).
6. **Data Visibility:** Datasets are visible only to users in the corresponding challenge. This is controlled in the DatasetController (DatasetController and ChallengeService).

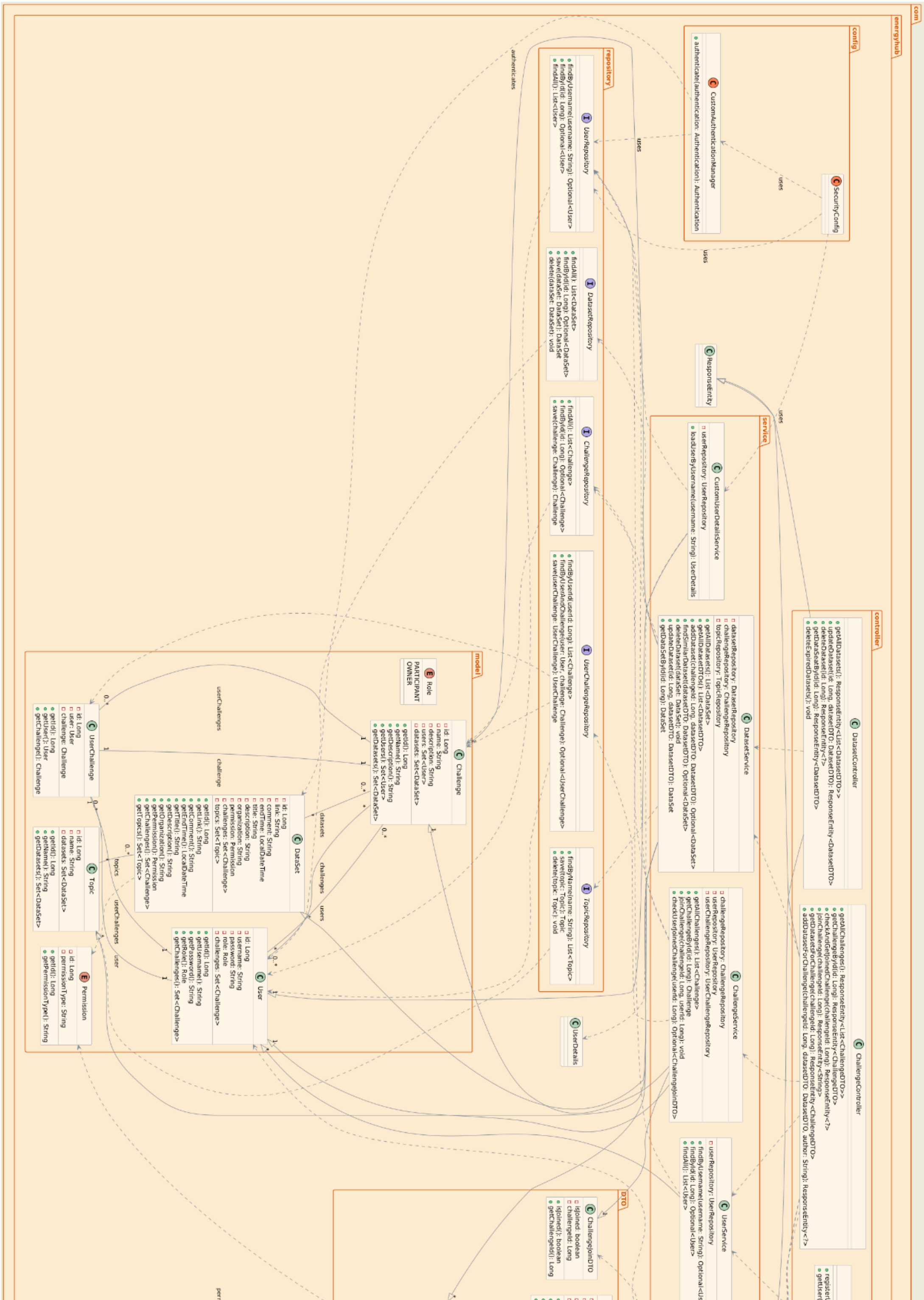


Figure 6: Backend architecture

3.2.4 Database

The database schema, hosted on Azure and managed by Liquibase, has evolved through several stages to support the application's data model.

Core Entities:
users: Stores user credentials (id, username, password) and roles (role). username is unique.
challenges: Stores information about challenges (id, name, description, tag).
datasets: Stores details about datasets (id, title, description, comment, link, organization, permission, end_time). An initial topic column was later removed.
topics: Stores categories for datasets (id, name), with name being unique.

Relationships:
challenge_datasets: A join table establishing a many-to-many relationship between challenges and datasets.
user_challenge: A join table establishing a many-to-many relationship between users and challenges.
dataset_topics: A join table establishing a many-to-many relationship between datasets and topics.

This schema, residing within an Azure database service, supports user management, challenge definition, dataset storage with associated metadata, and the categorization of datasets into topics, with relationships managed through dedicated join tables

Figure 7: DB tables schema

3.3 Testing

This chapter details the methodology employed to gather feedback on the developed high-fidelity prototype using surveys and questionnaires. Due to the focus on evaluating the user experience and perceived utility of the data management functionalities (without visualisations at this stage), surveys and questionnaires were deemed the most appropriate methods to efficiently collect feedback from a representative group of potential users (Nielsen, 1993).

The primary goal of this testing phase was to understand user perceptions regarding the accessibility and manageability of data within the prototype, even in the absence of visual representations. The survey instrument was designed to elicit both quantitative and qualitative data, allowing for a comprehensive understanding of user attitudes and identifying potential areas for improvement in the subsequent development stages, particularly concerning data visualisation (Sauro & Dumas, 2009).

3.6.1 Instrument Design

The survey combined Likert scale and open-ended questions, following best practices (Dillman et al., 2014).

- **Likert Scale Questions** measured user perceptions of data accessibility and management, asking participants to rate statements like:
    - "I found it easy to navigate to the data management section."
    - "The process for [specific task] was intuitive."
    - "I felt confident managing the data."
    - "Labels and instructions were clear."
      (Tullis & Albert, 2013)
- **Open-Ended Questions** allowed participants to provide detailed feedback, share challenges, and suggest improvements (Goodman et al., 2012), with prompts like:
    - "What aspects of data management were most confusing?"
    - "What features would you like to see added?"
    - "Other comments or suggestions?"

The survey focused on the current prototype's data management features (excluding visualisations) and used clear, user-friendly language (Krug, 2014).

In order to assess how users interact with a product and to evaluate its usability, specific methods are employed. The main focus lies on understanding how effectively users complete predefined tasks and identifying any difficulties they encounter along the way (Cooper et al., 2007). The key objectives of this process include:

1. **Assessing usability**: Evaluating how efficiently and successfully users accomplish tasks.
2. **Identifying issues**: Pinpointing areas where users face comprehension challenges.
3. **Discovering potential improvements**: Highlighting aspects of the design that could be enhanced.

3.6.2 Procedure for Usability Testing – Quantitative Evaluation

For assessing the final prototypes as completed products, a **quantitative testing strategy** was implemented. The purpose was to measure user satisfaction across different user groups. Participants completed an online survey composed of multiple-choice questions, aiming to determine how well the prototypes aligned with user expectations.
The test group consisted of both healthcare professionals and other relevant stakeholders. By including a diverse participant pool, the study aimed to capture broad feedback and potential differences in perception and use of the prototypes.

Testing Preparation, Execution, and Results

1. **Preparation and Prerequisites**:
    - High-fidelity prototype versions were set up and running locally.
    - Test participants were selected and confirmed.
    - Questionnaires and task descriptions were prepared by the project team and were ready for use during the test scenarios.
2. **Execution**:
    - The tasks that needed to be completed were clearly described to the participants.
    - While performing the tasks, participants simultaneously answered related questions.
    - All responses were collected and subsequently reviewed.
3. **Results**:
    - The number of data upload errors was recorded.
    - The number of clicks required to locate a specific dataset was measured.

General Experience

1. **Have you previously participated in a hackathon?**
    - Yes
    - No

Design and Flow

2. **How would you rate the overall design of the prototype?**
    - Very Good
    - Good
    - Neutral
    - Poor
    - Very Poor
    - Comments (optional): __________
3. **How would you rate the navigation and flow of the prototype?**
    - Very Intuitive
    - Intuitive
    - Neutral
    - Confusing
    - Very Confusing
    - Comments (optional): __________

Task-Based Questions

4. **Challenge 1** – Were you able to edit a dataset?
    - Yes, easily
    - Yes, but it was difficult
    - No, I could not figure it out
    - Comments (optional): __________
5. **Challenge 2** – Were you able to delete a dataset?
    - Yes, easily
    - Yes, but it was difficult

Figure 8: User Test part I

- No, I could not figure it out
- Comments (optional): __________

6. **Challenge 2** – Were you able to add a new dataset?
   - Yes, easily
   - Yes, but it was difficult
   - No, I could not figure it out
   - Comments (optional): __________
7. **Was it clear how to add a past dataset?**
   - Very clear
   - Clear
   - Neutral
   - Unclear
   - Very unclear
   - Comments (optional): __________

Feedback

8. **What did you like most about the Datahub?**

---

9. **What did you dislike about the Datahub?**

---

10. **What improvements would you suggest?**

Figure 9: : User Test part II

4 Result

This chapter details the outcomes of the developed prototype and the conducted tests. It begins by presenting the prototype created collaboratively with clients after analysis the needs and adapting depending on the clients feeback.. Subsequently, it analyzes the test results of the hi-fi prototype to assess the usability and understandability of different data setss management.

4.1 Lo-Fi Prototyping

The lo-fi prototype was developed in partnership with the commissioning parties through a meeting. The pros and cons of data handling and visualization were identified, leading to the creation of the prototype. The key findings are summarized in the next chapter.

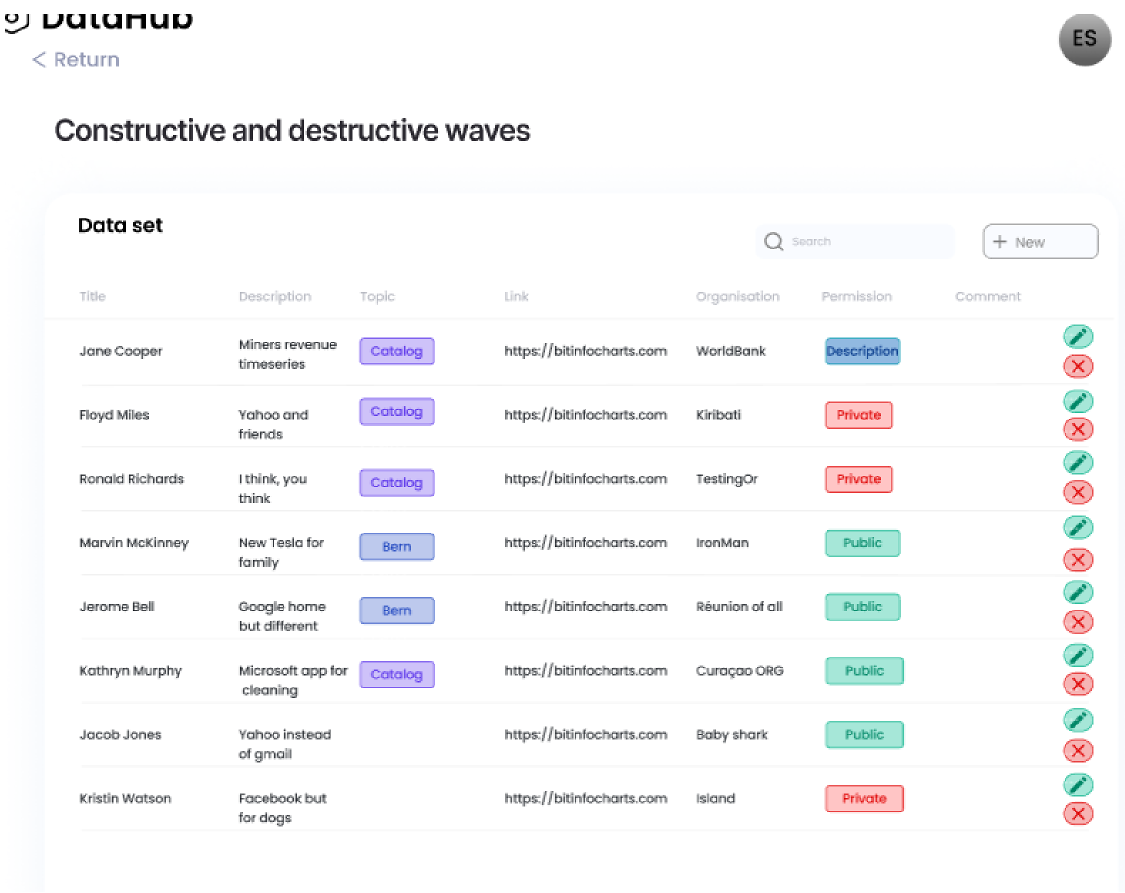


Figure 10: Data sets management

4.1.1 Data Management

The dataset should be easily manageable through the owner portal, allowing efficient handling and organization of the data.

4.1.2 Data Storage with Retention Time

A mechanism should be implemented to store the data for a defined period (dead time) before it is deleted or archived.

4.1.3 Form for Easy Input

A structured form should be provided within the portal to assist owners in easily adding new data entries.

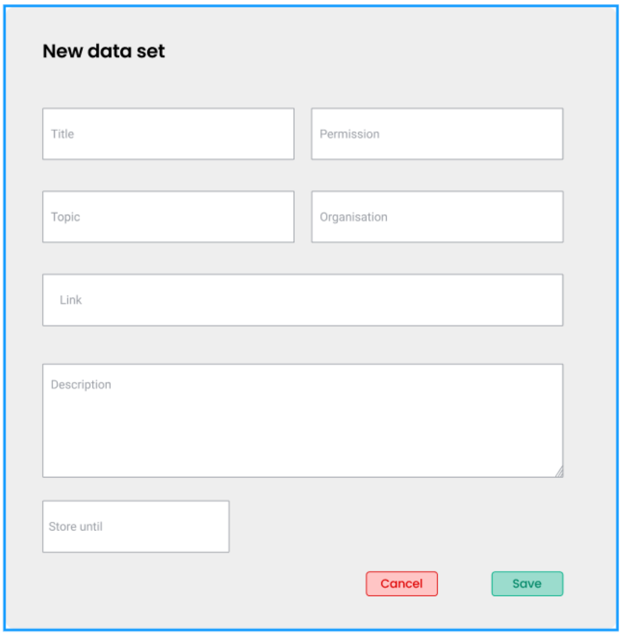
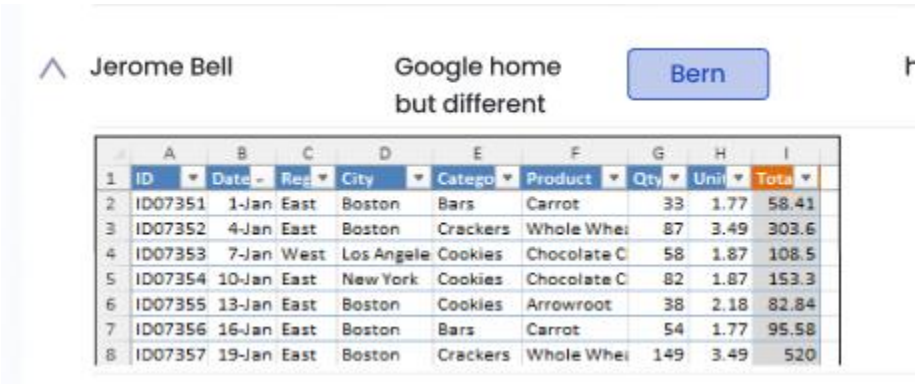


Figure 11: New data sets form

4.1.4 Participant Data Viewing

Participants should be able to view the dataset in a structured table format within the portal. All the data sets are in a centralized place and not to be dispense to multiple pages.



ID	Date	Reg	City	Catego	Product	Qty	Unit	Tota
ID07351	1-Jan	East	Boston	Bars	Carrot	33	1.77	58.41
ID07352	4-Jan	East	Boston	Crackers	Whole Whe	87	3.49	303.6
ID07353	7-Jan	West	Los Angele	Cookies	Chocolate C	58	1.87	108.5
ID07354	10-Jan	East	New York	Cookies	Chocolate C	82	1.87	153.3
ID07355	13-Jan	East	Boston	Cookies	Arrowroot	38	2.18	82.84
ID07356	16-Jan	East	Boston	Bars	Carrot	54	1.77	95.58
ID07357	19-Jan	East	Boston	Crackers	Whole Whe	149	3.49	520

Figure 12: Participants view of data sets

4.2 Test Results

This chapter summarizes the outcomes of the three testing phases conducted. The subsequent sections detail the findings from the qualitative usability testing, involving both a general group of testers and the intended end-user groups.

Data Entry Observations:

During the testing, participants interacted with data entry functionalities. Key observations include:

- A total of four datasets were added to the system. Notably, two of these additions involved the reuse of existing data entries.
- From the 2 manually added, these entries contained some errors already present within the system.
- One instance of data entry involved an incorrect data link.
- Despite these specific errors, all submitted datasets adhered to the minimum mandatory data requirements.
- Two datasets were successfully edited without any errors encountered by the tests.
- The deletion of two datasets was also completed without any reported issues.

Search Efficiency:

The efficiency of the search functionality was also evaluated. Participants required between a minimum of three clicks and a maximum of six clicks to locate a specific target dataset within the system.

Key Feedback Points:

Several key feedback points emerged from the testing sessions:

- **Lack of Data Visualization:** Testers noted the absence of data visualization features. It's important to acknowledge that this feedback is based on a version of the system prior to the integration of these visualization capabilities.
- **Confusion in Permission Naming:** Testers expressed confusion regarding the terminology used for different permission levels within the system. The naming conventions for permissions were not immediately clear to the testers.
- **Easy Navigation:** Testers found the navigation intuitive, stating that accessing challenges and their respective datasets required only "few clicks." This indicates a user-friendly information architecture.
- **Clear Form Instructions:** The forms effectively communicated the necessary information to testers. As one tester noted, the "form helped to know which information were needed of them," reducing ambiguity and ensuring clarity in data input.
- **Standardized Data Collection:** The forms facilitated the collection of consistent data across all submissions. Testers reported that the forms "helped to have the same standard for each data sets," which is crucial for data analysis and comparison.

4.3 Hi-Fi Prototyping

This chapter details the implementation of the selected lo-fi prototypes into hi-fi prototypes. The lo-fi prototypes were developed into realistic variants using Java and Javascript and employed for usability testing with the user group.

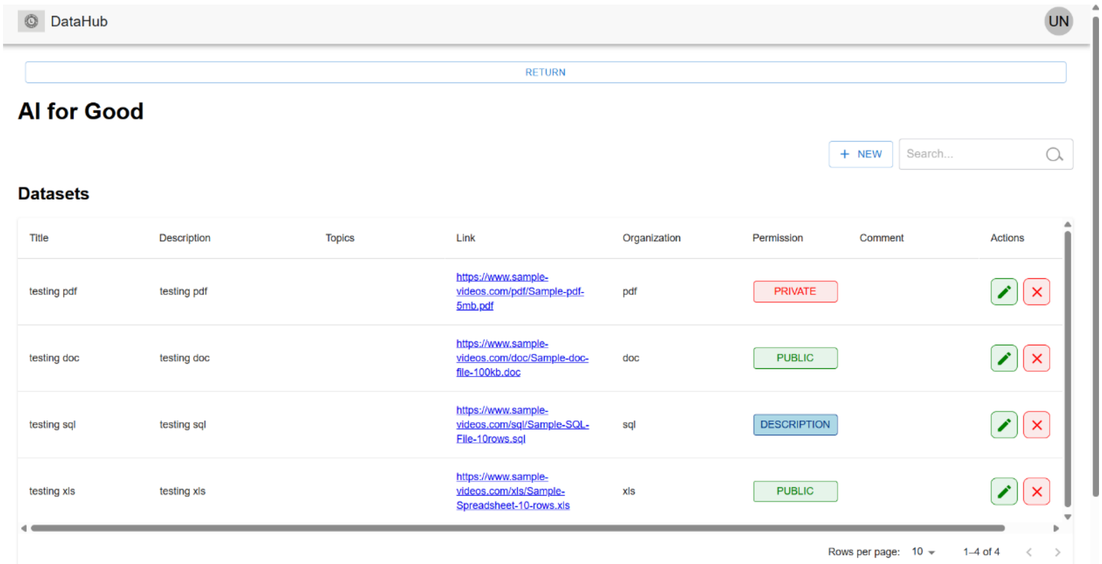


Figure 13: Owner data sets management

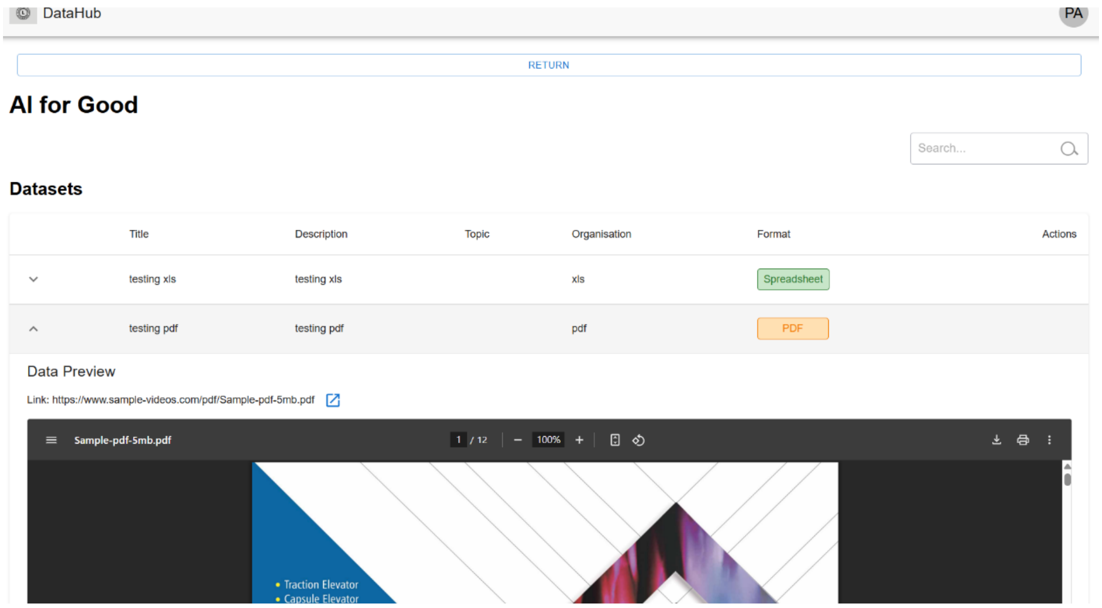


Figure 14: Participants data sets visualisation

4.3.1 Impact of Time Constraints on Prototype Evaluation

Time management challenges during the project led to an inability to conduct comprehensive testing of the prototype with its full functionality. These challenges, while specific to this project, underscore the critical importance of effective planning, realistic scheduling, and personal accountability in project management. The resulting limitation significantly reduced the opportunity to gather thorough user feedback necessary for adaptation and improvement, ultimately preventing the final hi-fi prototype from being officially passed by user testing.

4.4 Summary of result

Usability testing showed that while basic data entry worked, some errors and data reuse occurred. Editing and deletion were smooth. Searching took 3-6 clicks. Users found navigation intuitive and forms clear for standardized data collection, but noted a lack of visualization and confusing permission names in the tested version. Overall, the system was user-friendly and facilitated centralized data management through its forms.

5 Conclusion

This work began with the assumption that hackathons in the energy sector, play a crucial role in optimizing by innovative technologies solutions within the field. a platform for collecting and organizing data, was identified as particularly promising. The goal of this project was to develop effective methods for visualizing and managing the datasets provided by hackathon challenge owners, enabling participants to quickly and accurately assess the data for their challenges.

As part of this work, a solution was developed that delivers datasets from the energy field in a clear, simple, and standardized manner through a web application functioning as a data hub. Usability testing confirmed that this straightforward approach was particularly well-suited to the requirements of hackathons during events.

Additionally, accessing the datasets posed a particular challenge, as the data were often provided in different formats, making it difficult to process and manipulate them efficiently. Nevertheless, a solution was developed that addresses this issue, enabling a consistent and clear presentation of datasets. This consistency is crucial for ensuring the smooth, rapid organization and operation of hackathons.

In the future, there is significant potential for further development, particularly through the refinement and expansion of the current visualization methods. Future research should focus on testing the method developed in this work during real hackathon events to assess its long-term effectiveness and applicability. Additionally, exploring the standardization of datasets could enhance the accessibility and flexibility of visualizations, allowing for data manipulation through specific algorithms to generate visual results based on defined inputs. This approach could not only improve user-friendliness but also provide more opportunities for deeper engagement with the data during hackathons

References

Innosuisse. (2025, February 5). *The Innosuisse impact monitor*. https://www.innosuisse.admin.ch/en/impact-monitor

Pe-Than, E., & Herbsleb, J. D. (2016). *The hackathon phenomenon*. IEEE Software, *33*(5), 82-88.

Das, L., Saadat, M. H., Gjorgiev, B., & Sansavini, G. (2024, August 22). *Publication on advanced methods for the maintenance of the Swiss transmission grid*. ETH Zurich. Retrieved from https://rre.ethz.ch/the-lab/news/2024/08/publication-on-advanced-methods-for-the maintenance-of-the-swiss-transmission-grid.html

Bundesamt für Energie BFE. (2022, Dezember 16). *Energiestrategie 2050: Fünfjährliche Berichterstattung im Rahmen des Monitorings* https://www.bfe.admin.ch/bfe/en/home/versorgung/statistik-undgeodaten/monitoringenergiestrategie2050.exturl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWRtaW4uY2gvZGUvcHVibGljYX/Rpb24vZG93bmxvYWQvMTEyMzY=.html

Anderson, D. J. (2010). *Kanban: Successful evolutionary change for your technology business.* Blue Hole Press.

Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: The essentials of interaction design* (Completelyrev. and updated, [3. ed.]). Wiley.

Dam, R. F., & Siang, T. Y. (2024, Februar 21). *What is Design Thinking and Why Is It So Popular?* The Interaction Design Foundation. https://www.interaction-design.org/literature/article/what-is-designthinking-and-why-is-it-so-popular

Ferris, T. L. J. (2004). User-Centered Design: An Integrated Approach. *IEEE Transactions on Professional Communication*, *47*(1), 75–77. https://doi.org/10.1109/TPC.2004.824283

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, *53*(4), 50-58.

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, *25*(6), 599-616.

Dell Technologies. (n.d.). *PowerEdge Servers*. Retrieved from https://www.dell.com/en-us/shop/storage-servers-and-networking-for-business/sf/poweredge
Grossman, R. L. (2009). The case for cloud computing. *IT Professional*, *11*(2), 23-27.

Hayes, B. (2008). Cloud computing. *Computer*, *41*(7), 84-86.

Hewlett Packard Enterprise. (n.d.). *HPE ProLiant Servers*. Retrieved from https://www.hpe.com/emea_europe/en/hpe-proliant-servers.html

Hewlett Packard Enterprise. (n.d.). *HPE Synergy*. Retrieved from https://www.hpe.com/il/en/compute/synergy-infrastructure.html

Knorr, E. H., & Gruman, G. (2012). What cloud computing really means. *InfoWorld*, *1-12*.
Rappa, M. A. (2004). The utility business model and the future of computing services. *IBM Systems Journal*, *43*(1), 32-42.

Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, *34*(1), 1-11.
MellBuyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, *25*(6), 599-616.
Hurwitz, J., Bloor, R., Kaufman, M., & Halper, F. (2009). *Cloud computing for dummies*. John Wiley & Sons.
Knorr, E. H., & Gruman, G. (2012). What cloud computing really means. *InfoWorld*, *1-12*.
Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing* (Special Publication 800-145). National Institute of Standards and Technology.
Rimal, B. P., Choi, E., & Lumb, I. (2011). A taxonomy and survey of cloud computing systems. In *Ninth International Conference on Parallel and Distributed Systems* (pp. 667-674). IEEE.
Vaquero, L. M., & Buyya, R. (2011). Towards a cloud of clouds. *Future Generation Computer Systems*, *27*(5), 617-626.
Zhang, Y., Cheng, X., & Hu, L. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of Parallel and Distributed Computing*, *70*(11), 1179-1196.
Microsoft. (n.d. [a]). *Azure Data Factory documentation*. Retrieved from https://learn.microsoft.com/en-us/azure/data-factory/
Amazon Web Services, Inc. (n.d. [a]). *AWS Glue*. Retrieved from https://aws.amazon.com/glue/
Google Cloud. (n.d. [a]). *Dataflow documentation*. Retrieved from https://cloud.google.com/dataflow/docs
Microsoft. (n.d. [b]). *Azure Portal Overview*. Retrieved from https://learn.microsoft.com/enus/azure/azure-portal/azure-portal-overview
Amazon Web Services, Inc. (n.d. [b]). *Getting started with AWS*. Retrieved from https://aws.amazon.com/getting-started/
Google Cloud. (n.d. [b]). *Google Kubernetes Engine (GKE)*. Retrieved from https://cloud.google.com/kubernetes-engine
Microsoft. (n.d. [c]). *Azure Monitor documentation*. Retrieved from https://learn.microsoft.com/en-us/azure/azure-monitor/
Amazon Web Services, Inc. (n.d. [c]). *Amazon CloudWatch*. Retrieved from https://aws.amazon.com/cloudwatch/
Google Cloud. (n.d. [c]). *Google Cloud Operations suite (formerly Stackdriver)*. Retrieved from https://cloud.google.com/products/operations
Microsoft. (n.d. [d]). *Azure pricing calculator*. Retrieved from https://azure.microsoft.com/en-us/pricing/calculator/
Amazon Web Services, Inc. (n.d. [d]). *AWS Pricing*. Retrieved from https://aws.amazon.com/pricing/
Google Cloud. (n.d. [d]). *Google Cloud Pricing*. Retrieved from https://cloud.google.com/pricing
Microsoft. (n.d. [e]). *Azure Compliance Offerings*. Retrieved from https://learn.microsoft.com/en-us/microsoft-365/compliance/offering-home
Amazon Web Services, Inc. (n.d. [e]). *AWS Compliance Programs*. Retrieved from

https://aws.amazon.com/compliance/programs/
Google Cloud. (n.d. [e]). *Google Cloud Compliance Resource Center*. Retrieved from https://cloud.google.com/security/compliance
Microsoft. (2025). *Azure Arc Overview*. Retrieved from https://learn.microsoft.com/en-us/azure/azure-arc/
Amazon Web Services. (2025). *AWS Outposts Overview*. Retrieved from https://aws.amazon.com/outposts/
Microsoft. (2025). *Azure Arc Pricing*. Retrieved from https://azure.microsoft.com/en-us/pricing/details/azure-arc/
Amazon Web Services. (2025). *AWS Outposts Pricing*. Retrieved from https://aws.amazon.com/outposts/pricing/
Microsoft. (2025). *Use Cases for Azure Arc*. Retrieved from https://learn.microsoft.com/en-us/azure/azure-arc/overview
Amazon Web Services. (2025). *Use Cases for AWS Outposts*. Retrieved from https://aws.amazon.com/outposts/

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. National Institute of Standards and Technology.
Abramov, D., & Clark, A. (2015). *Building user interfaces with React*. Retrieved from https://facebook.github.io/react/
Eich, B. (1995). *The JavaScript language*. Netscape Communications Corporation. Retrieved from https://www.netscape.com
Facebook. (n.d.). *React – A JavaScript library for building user interfaces*. Retrieved from https://reactjs.org/
Google. (2014). *Material Design guidelines*. Retrieved from https://material.io/design
Material-UI. (n.d.). *Material UI - React components for faster and easier web development*. Retrieved from https://mui.com/
MDN Web Docs. (n.d.). *JavaScript*. Mozilla Developer Network. Retrieved from https://developer.mozilla.org/en-US/docs/Web/JavaScript
W3C. (2018). *Cascading Style Sheets (CSS) – A language for describing the presentation of Web pages*. World Wide Web Consortium (W3C). Retrieved from https://www.w3.org/Style/CSS/
Soares, M. M., Rosenzweig, E., & Marcus, A. (Hrsg.). (2022). *Design, User Experience, and Usability: UX Research, Design, and Assessment: 11th International Conference, DUXU 2022, Held as Part of the24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings, Part I* (Bd. 13321). Springer International Publishing. https://doi.org/10.1007/978-3-031-05897-4