Althena

D6.2 Report on data sharing and integration with European data lakes, OpenData and OpenTool

Dissemination level	Public (PU)
Work package	WP6
Task:	T6.4
Deliverable lead:	Vicomtech
Version	V1.0
Submission date	10/11/2025
Due date	31/10/2025

Authors

Authors in alphabetical order		
Name	**Organisation**	**Email**
Juan Diego Ortega	Vicomtech (VICOM)	jdortega@vicomtech.org

Control sheet

Version history			
Version	**Date**	**Modified by**	**Summary of changes**
0.1	20/10/2025	Juan Diego Ortega	Initial ToC and first contributions
1.0	10/11/2025	Juan Diego Ortega	Integration with CCAM TDS and final version

Peer review		
	Reviewer name	**Date**
Reviewer 1	Marcos Nieto (VICOM)	10/11/2025

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Disclaimer

This work has received funding from the Swiss State Secretariat for Education, Research, and Innovation (SERI).

Table of Contents

List of Figures

List of Tables

Glossary

Acronym	Definition
AI	Artificial Intelligence
CCAM	Connected, Cooperative and Automated Mobility
DFRS	Data for Road Safety
DMP	Data Management Plan
DOI	Digital Object Identifier
DSSC	Data Spaces Support Centre
EDC	Eclipse Data Space Connector
EMDS	European Mobility Data Space
FAIR	Findable, Accessible, Interoperable, and Reusable
FDS	Federated Data Sharing
FOT	Field Operational Test
GDPR	General Data Protection Regulation
IDSA	International Data Spaces Association
IDS-RAM	IDS Reference Architecture Model
ITS	Intelligent Transport Systems
KB	Knowledge Base
MLOps	Machine Learning Operations
NAP	National Access Point
NAPCORE	National Access Point Coordination Organisation for Europe
SRTI	Safety-Related Traffic Information
TDS	Test Data Space
XAI	Explainable Artificial Intelligence

1. INTRODUCTION

1.1. AIthena concept and approach

Connected, Cooperative and Automated Mobility (CCAM) solutions have emerged thanks to novel Artificial Intelligence (AI) which can be trained with huge amounts of data to produce driving functions with better-than-human performance under certain conditions. The race on AI keeps on building Hardware (HW) and software (SW) frameworks to manage and process even larger real and synthetic datasets to train increasingly accurate AI models. However, AI remains largely unexplored with respect to explainability (interpretability of model functioning), privacy preservation (exposure of sensitive data), ethics (bias and wanted/unwanted behaviour), and accountability (responsibilities of AI outputs). These features will establish the basis of trustworthy AI, as a novel paradigm to fully understand and trust AI in operation, while using it at its full capabilities for the benefit of society. AIthena will contribute to build Explainable AI (XAI) in CCAM development and testing frameworks, researching three main AI pillars: data (real/synthetic data management), models (data fusion, hybrid AI approaches), and testing (physical/virtual X- in the loop (XiL) set-ups with scalable Machine Learning Operations (MLOps)). A human- centric methodology will be created to derive trustworthy AI dimensions from user identified group needs in CCAM applications. AIthena will innovate proposing a set of Key Performance Indicators (KPI) on XAI, and an analysis to explore trade-offs between these dimensions. Demonstrators will show the AIthena methodology in four critical use cases: perception (what does the AI perceive, and why), situational awareness (what is the AI understanding about the current driving environment, including the driver state), decision (why a certain decision is taken), and traffic management (how transport-level applications interoperate with AI-enabled systems operating at vehicle level). Created data and tools will be made available via European data sharing initiatives (OpenData and OpenTools) to foster research on trustworthy AI for CCAM.

1.2. Purpose of this deliverable

This document is the final, public report for Task 6.4, "OpenData plan and OpenTool plan and sharing". The purpose of this deliverable is to document the analysis, strategy, and execution of the project's data sharing activities, ensuring that project results are made available to the wider European research community.

The scope of this report is twofold, directly addressing the objectives of Task 6.4:

1. **To Report on Analysis**: It presents a thorough analysis of the existing European data sharing landscape. This includes an investigation of European data lakes, OpenData platforms, OpenTool initiatives, and the major European Data Space frameworks relevant to the CCAM and AI domains.
2. **To Report on Action and Integration**: It details the concrete actions AITHENA has taken to integrate its project outputs, specifically the sample data generated in Task 2.1 and evaluation results from Work Package 5 (WP5), with these external initiatives. This includes a forward-looking plan for the integration of all final datasets.

1.3. Relationship to the AITHENA Data Management Plan (DMP)

This report is the operational *execution* and *public-facing summary* of the strategy defined in AITHENA's internal Data Management Plan (DMP). The DMP (submitted as deliverables D7.2, D7.3,

and the final D7.4 "Second update of the Data Management Plan") is a document that details the comprehensive inventory of all datasets, their handling, ethical considerations, and preservation.

As mandated by Task 6.4, the DMP serves as the foundational "guidelines for the selection of what data will be shared, defining access policies and licensing, according to... EC recommendations". Where the DMP is the project's internal *plan*, this D6.2 deliverable is the external *report* on its successful implementation, made public to ensure transparency and guide future researchers to the project's outputs.

1.4. The FAIR Data Principles as a Foundation for AITHENA's Sharing Strategy

AITHENA's data sharing strategy is built upon the FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles. The project's DMP explicitly states that all data "has been made to be Findable, Accessible, Interoperable, and Re-usable to the maximum extent possible". This deliverable demonstrates *how* this commitment was practically implemented. The FAIR principles, as detailed in Chapter 4 of the DMP, are the methodological foundation for the integration activities described in this report:

- **Findable:** Data is made findable by assigning globally unique and persistent identifiers (e.g., DOIs), describing it with "rich metadata," and indexing it in searchable resources. This is achieved by publishing metadata in the FAME CCAM Test Data Space (TDS) catalogue.
- **Accessible:** Data is made accessible via open and standardized protocols. AITHENA utilizes a hybrid approach, making public data accessible via its Zenodo Community and project website, while enabling federated access protocols for sensitive data via the FAME CCAM Test Data Space (TDS).
- **Interoperable:** Data interoperability is achieved by using standard, formal vocabulary and data-exchange standards "whenever possible".
- **Reusable:** Data is made reusable through clear licensing. The AITHENA DMP specifies the use of the **Creative Commons CC BY-NC-ND 4.0** license for its public uploads, ensuring legal clarity for future researchers.

1.5. Intended Audience

This deliverable is a public document intended for several key audiences involved in the future of European mobility and artificial intelligence. The primary audience is the European research and industrial community engaged in CCAM and Artificial Intelligence. This includes academic institutions, data scientists, and the R&D divisions of automotive manufacturers, technology suppliers, and software developers. For this audience, the report details what data is available, why it was created, and how to access it in accordance with FAIR principles.

A second key audience consists of stakeholders in the European data ecosystem, including public authorities, standardization bodies, and data space operators. The report’s analysis of the data sharing landscape, including the FAME CCAM Test Data Space, the European Mobility Data Space (EMDS), Gaia-X, and IDSA, provides a practical case study on implementing federated data sharing. For this group,

the deliverable serves as a reference for data governance, technical integration, and the application of the FAME Data Sharing Framework in a real-world project.

2. LANDSCAPE OF EUROPEAN DATA SHARING PLATFORMS

The European Union is actively building a single market for data, aiming to unlock its economic and societal value while ensuring data sovereignty, privacy, and fair competition. This strategy is built on a new legal framework, a federated infrastructure model (Data Spaces), and a strong commitment to open principles.

This chapter analyses the European data lakes, OpenData platforms, and OpenTool initiatives relevant to AITHENA's focus areas of CCAM, AI, and transport. This analysis informed the project's dissemination strategy, ensuring that AITHENA's outputs are directed to the most appropriate and impactful channels.

2.1. The Legal and Strategic Foundation

The EU's data strategy is anchored in two transformative pieces of legislation designed to govern the flow and use of data across the bloc.

- **The Data Governance Act (DGA)**, applicable since September 2023, focuses on establishing trust and creating robust mechanisms for data sharing. Its primary aim is to facilitate the re-use of specific categories of protected public sector data, such as confidential commercial data or health data, which cannot be released as simple OpenData. It does this by creating a legal framework for public bodies to authorize such re-use under strict conditions. Critically, the DGA introduces a novel regulatory framework for "data intermediation services", essentially neutral data marketplaces, which are forbidden from using the data they handle for their own profit, thereby ensuring they act as trusted brokers. Furthermore, it provides a common European framework for "data altruism," making it easier and safer for individuals and companies to voluntarily make their data available for the common good, such as for scientific research projects.
- **The Data Act**, set to become largely applicable from September 2025, directly addresses fairness and value allocation in the digital economy. A key provision grants both individuals and businesses the right to access the data generated by their connected systems, from smart appliances to industrial machinery. This empowers users to share this data with third-party service providers, for example, to obtain more competitive repair or maintenance services, breaking the data monopoly previously held by manufacturers. The Act also establishes rules for business-to-government (B2G) data sharing in cases of "exceptional need," such as public emergencies like floods or pandemics, compelling private companies to provide necessary data to public bodies. Finally, it introduces measures to make switching between cloud data-processing services easier, aiming to prevent vendor lock-in and foster a more competitive European cloud market.

2.2. Enabling Principles: OpenData, OpenTools, and Interoperability

The entire European data strategy is built upon a philosophy of openness, which is essential for ensuring interoperability and fostering innovation.

OpenData

This refers to the principle of making non-personal public sector data freely available for everyone to access, use, and share. It remains a foundational pillar, legally enshrined in the **Open Data Directive.**

- **Governing Law:** The **Open Data Directive** (which replaced the original Public Sector Information Directive).
- **Core Principle:** Public sector bodies should make their information reusable to foster transparency, innovation, and economic growth.
- **High-Value Datasets (HVDs):** A key concept from the directive. These are specific categories of data (Geospatial, Earth Observation, Meteorological, Statistics, Companies, and Mobility) that are considered so beneficial to society that they **must** be made available:
    - Free of charge
    - In machine-readable formats
    - Via Application Programming Interfaces (APIs) for easy access

OpenTool

While "OpenTool" is not an official EU term like "OpenData," it represents the underlying technical philosophy for building the data spaces. It embodies the principles required to make a federated system work:

- **Open-Source Software:** The EU's strategy encourages the use of open-source components for building data spaces, ensuring transparency, avoiding vendor lock-in, and fostering a collaborative development community.
- **Open Standards & APIs:** For data to flow between different organizations in a federated system, it must use common technical standards and open APIs. This is the only way to achieve **interoperability**, a cornerstone of the European Data Strategy.
- **FAIR Principles:** This concept, heavily promoted in EU-funded research (like Horizon Europe), is central to the "OpenTool" idea. All data, tools, and code should be:
    - **F**indable (e.g., with clear metadata)
    - **A**ccessible (e.g., via open protocols)
    - **I**nteroperable (e.g., using shared vocabularies)
    - **R**eusable (e.g., with clear licenses and source code)

2.3. Analysis of European Data Lakes and OpenData Initiatives

AITHENA's data outputs are designed to contribute to the growing body of European transport data. The following platforms represent key dissemination targets:

- **data.europa.eu:** This is the official, central portal for European open data, serving as the single point of access and aggregating datasets from 35 European countries. It features a vast and growing "Transport" category, which includes datasets on road networks, traffic, and public transport, driven by EU directives on public sector information and intelligent transport systems (ITS).
- **Data for Road Safety (DFRS):** This is a critical, industry- and public-authority-led ecosystem for operational safety data. Its purpose is to exchange live, near-real-time Safety-Related Traffic

Information (SRTI) using data from "anonymous probe vehicles and crowd-sourced data". This platform is vital for active safety services but is distinct from AITHENA's focus on R&D data.

- **The CCAM Knowledge Base (KB):** Maintained by the FAME project, the KB is the central resource hub for the European CCAM R&I community. It gathers data, knowledge, and experiences from R&I projects, providing a structured database of projects, methodologies (like the EU-CEM), and standards.

2.4. Overview of European OpenTool and Open-Source Initiatives

Beyond datasets, AITHENA is committed to sharing the *tools* and *models* developed within the project. The landscape for these outputs is equally rich:

- **Horizon Europe AI Initiatives:** The European Commission is actively fostering an open-source AI ecosystem through flagship initiatives such as GenAI4EU and OpenEuroLLM, which support the development of large-scale, open, and multilingual AI models. This creates a high-impact environment for sharing AITHENA's novel Explainable AI (XAI) models.
- **CCAM-Specific Hardware/Software Initiatives:** EC calls under Horizon Europe Cluster 4 and the Chips JU are specifically funding the development of optimized edge-AI algorithms and tools for CCAM systems. This indicates a strong, top-down-supported demand for the very type of XAI building blocks AITHENA is creating.
- **The 'Integrated CCAM Technologies Cluster':** This strategic alliance functions as a key collaborative "OpenTool" ecosystem, bringing together six EU-funded projects (AITHENA, CONDUCTOR, EVENTS, FRODDO, iEXODDUS, and PoDIUM). AITHENA's defined role within this cluster is to lead the definition of a "harmonized methodology for AI-based CCAM solutions". This cluster provides the primary, high-impact channel for sharing tools, models, and methodological synergies directly with peer R&I projects.

This analysis reveals that a sophisticated, multi-pronged dissemination strategy is necessary. AITHENA's different outputs (methodologies, software tools, and datasets) are not suited for a single platform. The project's strategy, therefore, directs each output to its most logical ecosystem:

1. **Methodologies** (e.g., D1.1, D5.1) contribute to the CCAM Knowledge Base.
2. **Tools and Software** (from WP3, WP4) are shared within the 'Integrated CCAM Technologies Cluster’.
3. **Datasets** (from T2.1) are integrated into the **FAME CCAM Test Data Space**, as detailed in the following chapters.

3. ANALYSIS OF EUROPEAN DATA SPACE INITIATIVES

This chapter analyses the foundational European Data Space initiatives. This analysis establishes the architectural and governance context for data sharing in Europe and concludes by identifying the specific gap for CCAM R&D data that AITHENA's strategy addresses.

3.1. Foundational Frameworks and Building Blocks

The European Strategy for Data is being realized through a set of interoperable frameworks and software building blocks that enable sovereign data sharing.

Gaia-X

Gaia-X is the high-level European initiative to build a federated, secure, and sovereign data infrastructure. Its mission is to create a *de facto standard* for data ecosystems based on European values (e.g., transparency, interoperability, and data sovereignty), allowing businesses and individuals to retain control over their data. Gaia-X provides the compliance and federation framework, not a physical infrastructure itself.

International Data Spaces Association (IDSA)

IDSA provides the *architectural blueprint* for implementing the Gaia-X vision: the **IDS Reference Architecture Model (IDS-RAM)**. The IDS-RAM is the standard for enabling trustworthy, self-determined data sharing. Its core component is the **IDS Connector**, which enforces data usage policies technically, allowing data providers to share data while ensuring it is only used according to a "technically enforced agreement".

Data Spaces Support Centre (DSSC)

The DSSC is an EC-funded initiative (under the Digital Europe Programme) designed to *support* and *harmonize* the creation of the various sectoral data spaces. It provides common requirements and best practices, offering key assets such as the **DSSC Blueprint** (a set of guidelines and a "starter kit") and the **DSSC Toolbox** (a catalogue of reusable software components and building blocks).

Simpl

Simpl is the EC's practical implementation of this vision. It is an open-source, secure *middleware* platform that enables data access and interoperability between data spaces. As described by the EC, "if DSSC is the city planner, Simpl is the infrastructure provider". Simpl provides the software to connect cloud-to-edge federations and allows data spaces to interoperate, re-using existing components like the Eclipse Dataspace Connector.

3.2. Mobility-Specific Data Space Implementations

These relevant frameworks are being applied to create a federated data ecosystem for the transport sector.

The Common European Mobility Data Space (EMDS)

The EMDS is the EC's flagship initiative to federate the multitude of existing transport data ecosystems and remove barriers to data sharing. It is being implemented by the deployEMDS project, which is

deploying an operational data space across nine implementation sites to accelerate smart, sustainable, and multimodal mobility.

NAPCORE (National Access Point Coordination Organisation for Europe)

NAPCORE is a critical, federated component of the EMDS. It coordinates and harmonizes the 30+ National Access Points (NAPs). These NAPs are mandated by the ITS Directive 2010/40/EU and are responsible for publishing operational mobility data, such as Safety-Related Traffic Information (SRTI) and C-ITS messages.

Eona-X

Eona-X is a non-profit, private-sector-led data space for mobility, transport, and tourism. Founded by major industry actors (e.g., Amadeus) , it is an endorsed Gaia-X project aiming to build an interoperable ecosystem for the travel and tourism industry.

3.3. Identified Gaps for CCAM R&D Data Sharing

There is a critical distinction and a clear gap in the European data space landscape. The major mobility data spaces initiatives (EMDS, NAPCORE, and Eona-X) are overwhelmingly focused on operational data to support mobility services. Their data assets consist of public transport schedules, traffic flow data, multimodal routing information, C-ITS messages, and SRTI.

The data generated by AITHENA, in the context of CCAM systems, is fundamentally different. It consists of:

- High-volume (e.g., 78 GB) raw sensor data (LiDAR, Radar, Camera).
- Specialized AI/ML training and validation datasets (e.g., V2AIX, MultiCorrupt).
- Robustness and uncertainty quantification datasets (e.g., OCCUQ).
- Scenario files for testing and validation.

This R&D data is of immense value to the CCAM research community and for the development of new AI models, but it is not directly applicable to operational mobility services. Placing this data in the EMDS would be an inefficient and inappropriate use of that platform.

This gap for CCAM-specific R&D data has been recognized by the CCAM Partnership. To fulfill this limitation, the FAME CCAM Test Data Space (TDS) has been developed. This specialized, federated platform is explicitly designed to facilitate the exchange of research and test data across R&I projects, making it the ideal and intended integration point for AITHENA's outputs.

4. APPLICATION OF THE FAME DATA SHARING FRAMEWORK IN AITHENA

This chapter analyses how the AITHENA project has applied the FAME CCAM Data Sharing Framework (DSF) 2.0[1]. This alignment ensures AITHENA's data management practices are compliant with European-level best practices for the CCAM community.

4.1. The FAME CCAM Data Sharing Framework (DSF)

The DSF is the guidance document for sharing data from CCAM testing, Field Operational Tests (FOTs), and Naturalistic Driving Studies (NDSs). As an output of the FAME support action, its purpose is to promote harmonization, accessibility, and interoperability of test data.

The framework provides comprehensive recommendations on all aspects of data sharing, including:

- **Data Sharing Agreements:** Guidance on Consortium Agreements, participant consent, and Data Space Actor Agreements.
- **Data and Metadata:** A rich model for categorizing and describing CCAM-specific data (Context, Acquired/Derived, Streaming, Aggregated) and metadata (Descriptive, Structural, Administrative).
- **Data Protection:** Detailed guidelines on GDPR compliance, anonymization, and security protocols.
- **Data Governance:** Formal procedures for data access, and a governance model for federated data spaces, including "Onboarding" and "Offboarding" procedures.

Crucially, the DSF 2.0 encourage **Federated Data Sharing (FDS)** as the primary approach. This model allows data providers to retain control and sovereignty over their data, granting access via trusted connectors and policies rather than traditional, centralized data transfers. As a result of the FAME project, the CCAM Test Data Space (TDS) was implemented to allow the R&D CCAM projects to offer and share the data which was generated during the time of the projects.

4.2. Comparative Analysis: AITHENA DMP alignment with the FAME DSF

The AITHENA project has taken the FAME DSF and aligned its principles with the foundational methodology for its own Data Management Plan.

The following table provides a direct, comparative analysis, mapping the core recommendations of the FAME DSF to their specific implementation within the AITHENA DMP. This analysis demonstrates AITHENA's full compliance and application of the framework.

[1] https://www.connectedautomateddriving.eu/data-sharing/ccam-data-sharing-framework/

FAME DSF Recommendation Area	Specific FAME DSF Guidance	AITHENA DMP Implementation	Analysis of Alignment
FAIR Principles	Recommends FAIR principles as the guidelines for improving data reusability	The DMP is structured around the FAIR principles, with Chapter 4 dedicated to their implementation.	**Full Alignment.** AITHENA's DMP is built on the exact FAIR principles advocated by the DSF.
Findability	Recommends data be "described with rich metadata" and assigned "globally unique and persistent identifier[s]"	DMP mandates "rich metadata," "standard identification mechanisms, such as Digital Object Identifiers" (DOIs), and a clear naming/versioning convention.	**Direct Implementation.**
Accessibility	Recommends "Access and authorisation need to be clearly defined" using "standardized communication protocols"	DMP specifies access via a dedicated AITHENA Zenodo Community and the project website, with clear protocols for anonymized data.	**Direct Implementation.**
Interoperability	Recommends (meta)data use "open, formal, standardized language, vocabularies, formats"	DMP states "Existing data-exchange standards are used whenever possible" and "(meta)data use vocabularies that follow FAIR principles".	**Direct Implementation.**
Reusability	Recommends "Rich metadata descriptions," "well-documented data provenance," and setting "a licence for data"	DMP mandates data deposition in Zenodo and other data sharing platforms with a specific, open license: **CC BY-NC-ND 4.0**.	**Direct Implementation.**
Data Governance (DMP)	Recommends the creation of a "Data Management Plan (DMP)" as a core governance document to collect relevant information	The AITHENA project's D7.4 is the "Second update of the Data Management Plan," a key deliverable of WP7.	**Full Alignment.** AITHENA treats the DMP as a living document, as recommended.
Data Protection (GDPR)	Provides extensive recommendations for data protection, privacy preservation, anonymization, and GDPR compliance.	The AITHENA DMP has dedicated chapters for "Data Security" and "Ethical Aspects", which details "Compliance with GDPR".	**Full Alignment.**

5. AITHENA'S INTEGRATION WITH THE FAME CCAM TEST DATA SPACE (TDS)

This chapter documents the practical integration actions AITHENA has completed in accordance with the user query. It describes the selected FAME CCAM TDS platform and details the successful sharing of initial datasets, which served as a proof-of-concept for the final integration plan in Chapter 6.

5.1. The FAME CCAM TDS: A Federated Solution for CCAM Research

As established in Section 3.3, the FAME CCAM TDS is the platform that has emerged to fill the data sharing gap for CCAM R&D projects. It is a key output of the FAME project, developed to demonstrate the concept of *federated data sharing* based on the principles of the DSF 2.0.

The architecture of the TDS is its most important feature. It is not a centralized "data lake" where data is uploaded and stored. Instead, it is a federated data space demonstrator that implements a **"connector-as-a-service"** approach. It utilizes open-source components, specifically the Eclipse Data Space Connector (EDC), to achieve this.

For data providers like AITHENA, this means:

1. **Data Sovereignty:** The data (e.g., large sensor files) remains on the provider's own servers.
2. **Discoverability:** A web-based dashboard allows the provider to *publish metadata* about their datasets, making them searchable in a central catalogue.
3. **Controlled Access:** The provider can set conditions for data access using policies. Data users can handle *semi-automatic contracts* via the dashboard, which, upon approval, grants them access via the secure connector.

This model is the one valid solution for sharing the type of large-scale, sensitive, and IPR-heavy datasets generated in CCAM research. The architecture of the CCAM TDS is depicted in the following figure.

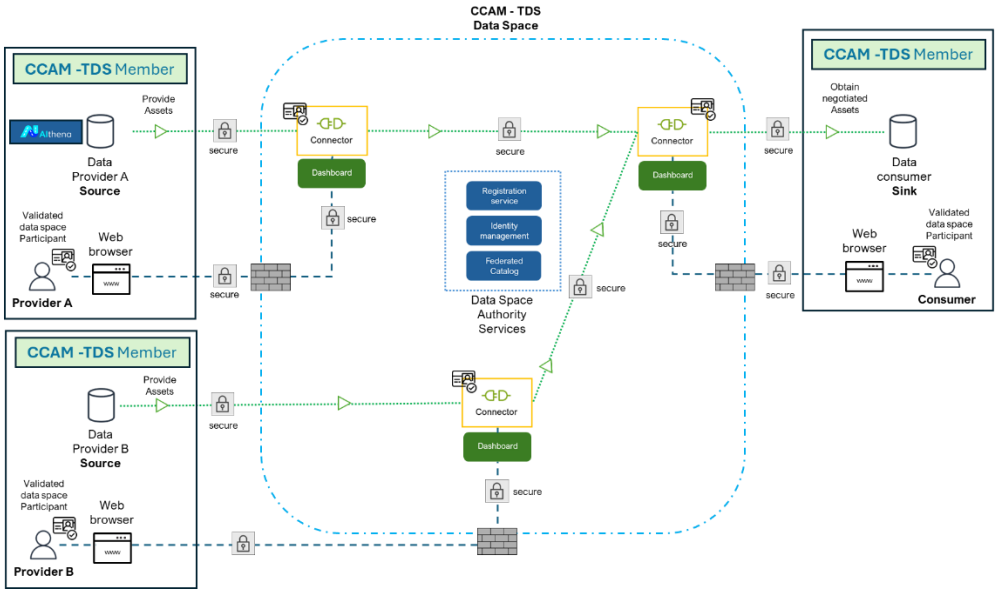

Figure 1: CCAM Test Data Space architecture and AITHENA's connector implementation.

5.2. AITHENA's Onboarding Process in CCAM TDS

The onboarding of the AITHENA project onto the FAME CCAM TDS involves following the onboarding procedure stablished by the FAME CCAM TDS and the principles outlined in the FAME DSF.

The onboarding process followed the Data Space Governance model outlined in the FAME DSF. Vicomtech, acting as the AITHENA coordinator, initiated the “onboarding” in the CCAM TDS. This process allowed AITHENA to:

1. Seamlessly register as a "Data Provider" within the TDS identity hub.
2. Collaborate on the definition of the metadata schema for the TDS catalogue, ensuring it could capture the specific attributes of AITHENA's R&D data.
3. Provision and test the Eclipse Data Space Connector (EDC) instance for AITHENA's data-hosting servers.
4. Develop and test the "semi-automatic contract" policies for AITHENA's private datasets.

This process served as the initial validation of the TDS itself, establishing a best-practice template for other CCAM Partnership projects to follow.

5.3. Sharing of Initial Datasets

As part of this initial onboarding and testing phase, AITHENA shared test datasets on the FAME TDS. This exercise was crucial for validating the federated access model and demonstrate the capabilities of sharing the generated or used data.

The DMD Dataset

- **Dataset:** The Driver Monitoring Dataset (DMD), a large, multi-modal dataset (RGB, Depth, IR) of real driving actions, including distraction, fatigue, and gaze allocation.
- **Relevance:** The DMD is critical for training and validating Driver Monitoring Systems (DMS), which are essential for safety in SAE Level 2 and Level 3 automated driving—a core topic for AITHENA.
- **"(Private)" Status Explained:** To exemplify sharing of private data. The DMD contains sensitive footage of real drivers. Although the dataset is open in other platforms, in the CCAM TDS it requires to follow a process of verification of policies between the consumer and the AITHENA’s connector.
- **TDS Integration:** The dataset was made discoverable on the FAME TDS, meaning its metadata is public in the catalogue. However, to access the data, a user must request it via the TDS dashboard and agree to the sharing contract. This triggers a contract negotiation workflow between AITHENA’s Data Space Connector and the consumer connector. This successful integration demonstrated the TDS's core capability, enabling discoverability while enforcing data sovereignty and controlled access for sensitive R&D data.

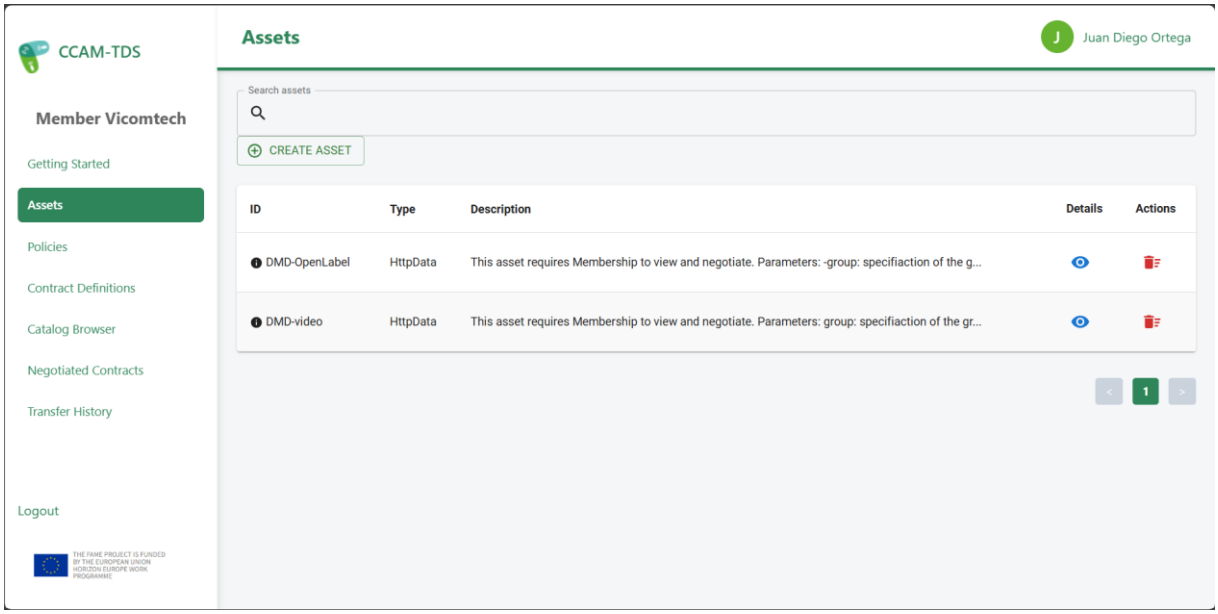

Figure 2: DMD datasets (video and OpenLabel services) added as data assets to CCAM-TDS.

The Nuscene (Private) Dataset

- **Dataset:** The Nuscene dataset, a well-known, large-scale public dataset for autonomous driving, which was re-used by AITHENA for its research.
- **Relevance:** AITHENA's research (e.g., in WP3) generated new findings, models, and potentially new annotations based on the Nuscene dataset.
- **"(Private)" Status Explained:** In this context, "Private" does not refer to the original Nuscene data, which is public. It refers to AITHENA's proprietary, value-added fork or annotated version of the dataset.
- **TDS Integration:** AITHENA registered its specific version of the Nuscene data in the FAME TDS. This registration, again with a "Private" access policy, protects the project's intellectual property and preliminary results during its runtime. This case study demonstrated the TDS's flexibility in managing not just new raw data, but also derived and modified data assets, allowing the project to "pre-register" its outputs securely.

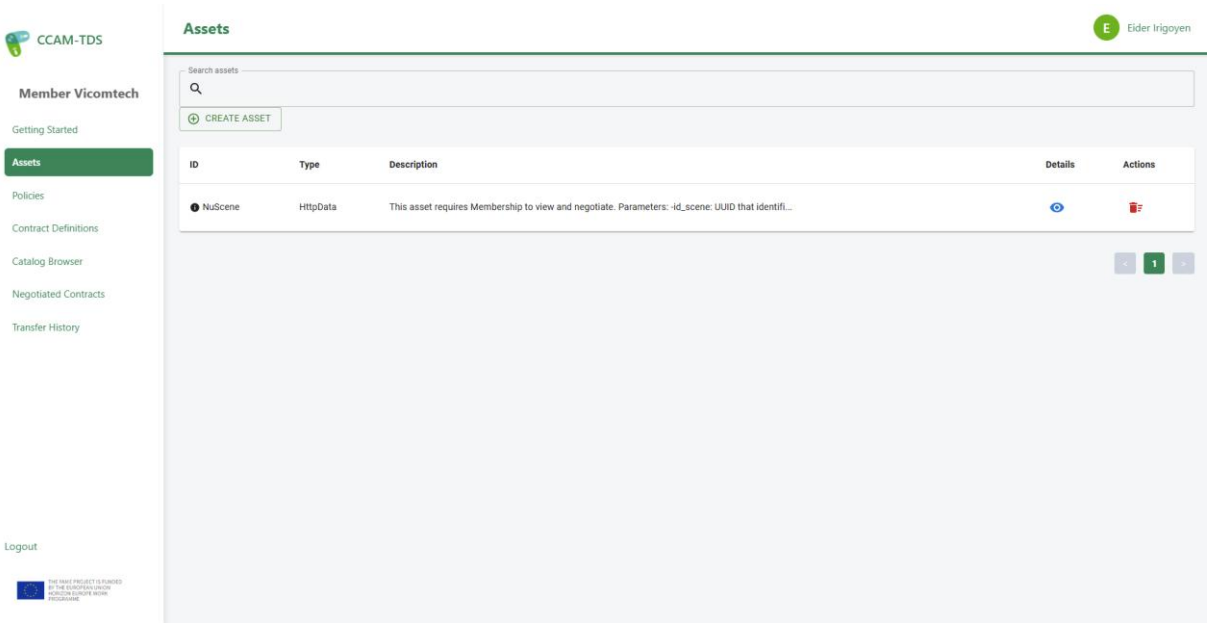

Figure 3: NuScene dataset added as data assets to CCAM-TDS.

6. PLAN FOR INTEGRATION OF AITHENA-GENERATED DATASETS

This chapter presents the forward-looking, actionable plan for the final integration of all novel datasets generated by AITHENA. This plan fulfills the primary objective of Task 6.4 by providing a plan to integrate the generated datasets into the FAME Test Data Space, according to the FAME Data Sharing Framework.

6.1. Inventory of AITHENA Datasets for FAME TDS Integration

The final inventory of novel datasets generated by AITHENA (originating from T2.1) is designated for sharing via the FAME TDS. The following plan details the specific integration status for each dataset, sourced from the project's final Data Management Plan.

Dataset Name	Description & Format	Data Category (per FAME DSF)	AITHENA FAIR Access Policy	Intended FAME TDS Integration Status
Siemens real-world dataset	Dataset in MCAP format containing LiDAR, Radar, and GNSS/IMU sensor data... from urban areas in Leuven... approximately 78 GB.	Acquired Data (Raw Sensor Data)	No (Internal AITHENA)	**Federated Access:** Metadata to be published on FAME TDS. Data remains on AITHENA (VICOM) server. Access requires contract negotiation due to large size and potential sensitivity.
V2AIX	A Multi-Modal Real-World Dataset of ETSI V2X ITS Messages in Road Public Traffic.	Acquired Data (Streaming Data, C-ITS)	Yes (Publicly Accessible URL: https://v2aix.ika.rwth-aachen.de/)	**Discoverable:** Metadata to be published on FAME TDS. The "access" link will point directly to the existing public URL.
MultiCorrupt	A Multi-Modal Robustness Dataset and Benchmark of LiDAR-Camera Fusion for 3D Object Detection.	Acquired/Derived Data (Sensor/ Benchmark Data)	Yes (Public GitHub URL: https://github.com/ika-rwth-aachen/MultiCorrupt)	**Discoverable:** Metadata to be published on FAME TDS. The "access" link will point to the public GitHub repository.
OCCUQ	Efficient Uncertainty Quantification for 3D Occupancy Prediction.	Derived Data (Processed AI Dataset)	Yes (Public GitHub URL: https://github.comika-rwth-aachen/OCCUQ)	**Discoverable:** Metadata to be published on FAME TDS. The "access" link will point to the public GitHub repository.
AITHENA Synthetic dataset	Synthetic dataset, alongside a user	Generated Data (Simulation/ Digital Twin)	Yes (AITHENA Zenodo Community: htt	**Discoverable:** Metadata to be published on FAME TDS. The "access" link will point directly

	manual, a white paper and an article stating the uses, benefits and drawbacks.	ps://zenodo.org/records/16680404)	to the dataset's DOI on the AITHENA Zenodo community.

Table 1: AITHENA-Generated Datasets for Integration into FAME TDS

6.2. Data Governance, Licensing, and Access Policy

This plan is governed by the principles of the FAME DSF.

- **Governance:** The integration will follow the "Data Space Governance" procedures. AITHENA, as the "Data Provider," will manage its data assets through the FAME TDS dashboard.
- **Licensing:** As defined in the AITHENA DMP and aligned with FAME DSF recommendations , all publicly shared datasets (V2AIX, MultiCorrupt, OCCUQ, and the AITHENA Synthetic dataset) are or will be licensed under a **CC BY-NC-ND 4.0** license. This license will be embedded in the FAME TDS metadata to ensure legal reusability.
- **Access Policy:** The plan utilizes a hybrid access model, demonstrating the full capability of the FAME TDS:
    - **Open Access (Discoverable):** For the four datasets already public, the FAME TDS will serve as a central *catalogue*. It will ingest their metadata, making them discoverable to the CCAM community and providing a direct link (via URL or DOI) to their open-access locations (GitHub, Zenodo, etc.).
    - **Restricted Access (Federated):** For the "Siemens real-world dataset," which is large and consortium-private, the FDS model will be fully implemented. The data will *remain* on AITHENA's servers. The FAME TDS will host the metadata. A researcher discovering the dataset can request access, which will trigger an approval workflow. This demonstrates how to share sensitive, large-scale R&D data responsibly.

6.3. Post-Project Onboarding Plan and Partner Responsibilities

As the AITHENA project has concluded, this deliverable serves as the final report and a guide for the post-project actions required to complete the integration plan. The steps defined during the project's lifetime now transition into a set of post-project recommendations for the relevant partners to ensure the AITHENA data legacy.

The following steps outline the actions each responsible partner must perform to publish their datasets in the FAME CCAM TDS, adhering to the FAME Data Sharing Framework.

Step 1: Formal Onboarding Initiation

The data-owning partners, IKA (for V2AIX, MultiCorrupt, OCCUQ) and SIE-BE (for the Siemens real-world dataset), are responsible for initiating the formal onboarding process outlined in the FAME DSF with the FAME CCAM TDS operator. This step involves registering as official "Data Providers" within the TDS and requesting a “Connector”.

Step 2: Metadata Finalization

IKA and SIE-BE must use the data inventoried in the final AITHENA Data Management Plan (D7.4) and this report to finalize the complete metadata descriptions for their respective datasets. This metadata must comply with the FAME DSF "Metadata documentation template" to ensure findability and interoperability. VICOM, will perform this action for the "AITHENA Synthetic dataset" hosted on Zenodo.

Step 3: Policy & Contract Definition

For the "Siemens real-world dataset," which follows a "Federated Access" model (see Table 6.1), partner SIE-BE will define the Data Space policies their datasets will have. These policies will govern how external researchers request and are granted access to the data which remains on SIE-BE's servers.

Step 4: Technical Integration of Datasets

For Open Access ("Discoverable") Datasets: IKA and VICOM will provide the finalized metadata and public-access URLs/DOIs for the V2AIX, MultiCorrupt, OCCUQ, and AITHENA Synthetic datasets during the creation of a new data asset in the CCAM TDS. Policies will be attached, and new data contracts are created, populating the catalogue of the CCAM TDS.

For Restricted ("Federated") Datasets: SIE-BE will work with the TDS operator (VICOM) to technically configure the secure connection between the FAME TDS Eclipse Data Space Connector (EDC) and the SIE-BE server (or designated VICOM-hosted server) hosting the "Siemens real-world dataset." This action is the core of the federated "connector-as-a-service" model , enabling data sharing without data transfer.

7. CONCLUSIONS

This report has detailed the successful execution of AITHENA's data sharing strategy, a core component of its dissemination and impact-generation activities. The project has moved from analysis to concrete implementation, navigating the complex European data ecosystem to find the most effective channels for its outputs.

The project's key achievements in this area are:

- **Methodological Alignment:** AITHENA has developed its internal Data Management Plan in full alignment with the European-level best practices of the FAME CCAM Data Sharing Framework 2.0.
- **Strategic Gap Analysis:** The project successfully analyzed the European data space landscape, identifying the critical gap between operational mobility platforms (EMDS, NAPCORE) and the specific needs of the CCAM R&D community.
- **Pioneering Platform Adoption:** AITHENA identified the FAME CCAM Test Data Space (TDS) as the correct solution and acted as a pioneer data provider. Currently the FAME CCAM TDS is the only data space devoted to share CCAM data.
- **Hybrid Sharing Model:** AITHENA has successfully implemented a sophisticated, hybrid sharing model. This includes sharing initial test data (DMD, Nuscene) and executing a final plan to share all generated datasets (V2AIX, MultiCorrupt, Siemens dataset, etc.) via a combination of open-access (discoverable) and federated (restricted) policies.

AITHENA's legacy extends beyond the development of trustworthy XAI models. Its lasting contribution is also the portfolio of valuable, FAIR-compliant, and highly specialized datasets for CCAM research.

By embracing a federated data sharing model via the FAME TDS, AITHENA has provided a best-practice template for future Horizon Europe projects. It has demonstrated how to responsibly share large-scale, sensitive R&D data while maintaining full data sovereignty. This work directly strengthens the CCAM Partnership, provides critical data to accelerate European research in trustworthy AI, and ensures that the project's results remain a durable asset for the community long after the project's conclusion.