D.3.3 Report on data management for applications in CECC – Initial

Document Summary Information

Project Identifier	HORIZON-CL4-2022-DATA-01. Project 101093129		
Project name	Agile and Cognitive Cloud-edge Continuum management		
Acronym	AC^3		
Start Date	January 1, 2023	**End Date**	December 31, 2025
Project URL	www.ac3-project.eu		
Deliverable	D3.3 - Report on data management for applications in CECC – Initial		
Work Package	WP3		
Contractual due date	30/06/2024	**Actual submission date**	28/06/2024
Type		**Dissemination Level**	PU
Lead Beneficiary	UBI		
Responsible Author	Dimitris Klonidis (UBI)		
Contributors	D. Amaxilatis (SPA), N. Tsironis (SPA), S. Sengupta (ION)		
Peer reviewer(s)	A. Ksentini (EUR), C. Verikoukis (ISI)		

Funded by the European Union

AC^3 project has received funding from European Union's Horizon Europe research and innovation programme under Grand Agreement No 101093129.

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
v0.1	22/02/2024	0%	Initial Deliverable Structure	D. Klonidis (UBI)
v0.2	09/05/2024	15%	Revisited structure, initial contents, agreed contributions per partner	D. Klonidis (UBI), N. Tsironis, D. Amaxilatis (SPA), S. Sengupta (ION)
v0.3	16/05/2024	10%	Editing of Section 2	N. Tsironis (SPA), D. Klonidis (UBI)
v0.4	17/05/2024	40%	Editing of Section 3	D. Amaxilatis (SPA)
v0.5	20/05/2024	60%	Editing of Section 4	S. Sengupta (ION) D. Amaxilatis (SPA)
v0.6	25/05/2024	75%	Editing of Section 5	N. Tsironis (SPA)
v0.7	07/06/2024	95%	Final editing and corrections	D. Klonidis (UBI)
v0.8	20/06/2024	80%	Internal Review	A. Ksentini (EUR)
v0.9	21/06/2024	90%	Updates and Executive Summary	D. Klonidis (UBI)
v1.0	30/06/2024	100%	Final Review and Submission	C. Verikoukis (ISI)

Disclaimer

Copyright message

Table of Contents

Executive Summary 7
1 Introduction 8
1.1 Overview – Purpose and objectives 8
1.2 Link with other project activities 8
1.3 Mapping AC^3 Outputs 9
1.4 Deliverable Overview and Report Structure 9
2 Data management in the overall architecture 10
2.1 Main concept and targeted functionalities 10
2.2 Mapping to the AC^3 architecture 11
2.3 Targeted development areas 14
2.3.1 Interfacing with End Users 14
2.3.2 Interfacing with Data Sources 14
2.3.3 Use of Data Sources through Service Deployment and Management 15
3 Data sharing and usage by end users 16
3.1 Overview of the implementation plan 16
3.2 Initial developments 17
3.2.1 Data Types of project Use Cases 18
3.2.2 Data catalogues and datasets in the AC^3 Data Catalogue 18
3.3 Planned implementations and extensions for the final phase 21
3.3.1 Linking Data Catalogue with the AC^3 Service Catalogue 21
3.3.2 Linking Data Catalogue to Semantic Aware Reasoner 21
3.3.3 Finalize the data sources and datasets of the project 22
3.3.4 Align the ontologies used in the Data Catalogue with the Service Catalogue 22
4 Interfacing with data sources and data collection mechanisms 23
4.1 Overview of the implementation plan 23
4.1.1 Adopted technologies 23
4.2 Initial developments 23
4.2.1 Data source registration & maintenance 23
4.2.2 Data Source Connectors 23
4.3 Planned implementations and extensions for final phase 26
5 Data management through service deployment and orchestration 27
5.1 Overview of implementation plan 27
5.1.1 Micro-service-based data management application design 27
5.1.2 Adopted technologies 28
5.2 Initial developments 28
5.2.1 Hot Data management application addons 28
5.2.2 Cold Data management application addons 29
5.3 Planned implementations and extensions for final phase 30
6 Conclusions 31

List of Figures

Figure 1 The updated AC3 Architecture with the data management framework elements being highlighted 12
Figure 2: Conceptual diagram of AC3 Data Space 13
Figure 3 The AC3 Data Catalogue 16
Figure 4: The components that form the AC3 Data Catalogue, their interactions, and interfaces 17
Figure 5: High Level View of an AC3 Data processing application 27
Figure 6: Structure of an AC3 Hot Data Application implementing a ML Service 28
Figure 8: Structure of an AC3 Cold Data Application processing stored Video or FITS files 29

List of Tables

Table 1: Adherence to AC3 GA Deliverable & Tasks Descriptions 9
Table 2 Description of the Data Types to be handled in the context of AC3 based on the Use Cases of the project 18

Glossary of terms and abbreviations used

Abbreviation / Term	Description
AC^3	Agile and Cognitive Cloud edge Continuum management
AG	Application Gateway
API	Application Programming Interface
AMQP	Advanced Message Queuing Protocol
CECC	Cloud Edge Computing Continuum
CECCM	Cloud Edge Computing Continuum Manager
CFN	Compute First Networking
CFRA	Cloud Federation Computing Reference Architecture
CNCF	Cloud Native Computing Foundation
CRUD	Create, Read, Update, and Delete
FHS	Federation hosting service
GUI	Graphical User Interface
IDSA	International Data Spaces Association
IoT	Internet of Things
KPI	Key Performance Indicator
LCM	Life-Cycle Management
LMS	Local Management System
ML	Machine Learning
NBI	Northbound Interface
OSR	Ontology and Semantic aware Reasoner
OWL	Web Ontology Language
PaaS	Platform as a Service
PLI	Profile Language Interpreter
RDF	Resource Description Framework
SBI	Southbound Interface

SIIF	Standard for Intercloud Interoperability and Federation
SLA	Service Level Agreement
UI	User Interface
WAN	Wide Area Network

Executive Summary

The main purpose of the work detailed in this deliverable (D3.3) is to establish the initial data management mechanisms within AC3, one of the key innovations of the project. Designed as a Platform-as-a-Service (PaaS) entity, AC3 incorporates advanced data management functionalities within its platform architecture to provide native support for all data-related aspects during the deployment and management of applications over the CEC infrastructure. This integration enables application developers to seamlessly incorporate data lifecycle management into their deployed applications, emphasizing efficient data integration, interoperability, dynamic data discovery, cataloguing, advanced processing capabilities, stringent security measures, scalability, and resilience. As a result, the AC3 data management framework positions itself as a comprehensive solution for managing data across diverse settings.

The deliverable focuses on three targeted development areas within the AC^3 architecture's data management capabilities: **(a)** interfacing with end users, ensuring intuitive data discovery and secure, compliant data sharing; **(b)** interfacing with data sources, enhancing adaptability and performance of Data Connectors to efficiently manage diverse data types and volumes; and **(c)** utilizing integrated data sources for dynamic service deployment and management, leveraging real-time insights to optimize resource usage, enhance service performance, and align with user requirements and system constraints for continuous improvement. More specifically:

The AC^3 Data Management Platform leverages the AC^3 Data Catalogue service for sharing data, providing comprehensive descriptions, metadata, and guidance to developers. This service ensures that data sources from various domains are well-described and annotated with rich semantic information based on established ontologies. Utilizing the European data management ecosystem Piveau, the AC^3 platform focuses on interoperability and open standards, offering components like the Piveau Hub UI, Hub Repository, Hub Search, Virtuoso RDF Triple Store, and Elasticsearch for robust data management and advanced search capabilities. Authentication and authorization are managed through Keycloak, a widely-used Open Source Identity and Access Management software integrated with Piveau, enhancing user interaction with the data catalogue.

With respect to the data source interfacing and collection mechanisms, the implementation is based on the Eclipse Data Space Connector framework, which aims to facilitate secure sharing and management of diverse data types, including IoT and video data, with functionalities such as real-time data streaming, efficient data routing, and robust data security measures. Integration with the Piveau Catalogue enhances metadata management and promotes open data compatibility, fostering better data cataloguing and utilization. Additionally, the introduction of the Cold Data Connector, leveraging IONOS S3 services, enhances the security, scalability, and reliability of data storage, crucial for the AC^3 project's objectives. These developments collectively lay the foundation for a robust, secure, and efficient data space within the AC^3 consortium, poised to drive innovation, collaboration, and data-driven decision-making across various sectors and stakeholders.

For the data management application and network service components, specific add-ons and data source connection guidelines have been established to support use case development. AC^3 applications employ a micro-service architecture, with these add-ons providing multiple microservices to meet application developers' needs. Key achievements include deploying components such as the Data Connector, Message Broker, Data Mapper, shared storage location, and one or more Data Manipulators. Technologies adopted are primarily based on microservices architecture, the AMQP protocol for communication, and shared storage solutions for data management. These micro-service add-ons are integrated into the AC^3 application to streamline data processing throughout the application's lifecycle. Planned implementations focus on leveraging real-time and historical data insights for dynamic service deployment and optimization, enhancing service management frameworks through data-driven decision-making, and enabling continuous service improvement. Future developments aim to expand data catalogue interfacing and incorporate additional data types to enhance end users' data sharing and utilization.

1 Introduction

This section provides an overview of the work reported in this deliverable, including the main purpose and the related key objectives. It also describes the link with the other project activities and related deliverables. The targeted outcomes are summarized next and mapped to the GA activities and the deliverable sections. The final subsection provides the structure of the deliverable.

1.1 Overview – Purpose and objectives

The purpose of the work presented in this deliverable (D3.3) is to provide details on the Initial data management mechanisms adopted in AC³ allowing application developers to incorporate the data life cycle management within the deployed applications.

The data management mechanism is designed as a Platform-as-a-Service (PaaS) entity and implemented with the goal of establishing efficient data integration and interoperability within a fluid environment. This involves seamlessly connecting heterogeneous data sources and sinks to enable smooth information flow across the system. Additionally, emphasis is placed on dynamic data discovery and cataloguing to facilitate intuitive access to resources, thereby simplifying data utilization. Furthermore, the data management framework prioritizes the support of advanced data processing and analysis capabilities, which is essential for real-time insights and informed decision-making in high-speed environments. Security remains a key focus, with stringent mechanisms for data exposure and sharing in place to ensure compliance with privacy standards. Lastly, the infrastructure is designed to be scalable and resilient, capable of accommodating evolving data volumes and processing demands without compromising performance or stability. These goals collectively position the AC³ data management framework as a comprehensive and robust solution for managing data across diverse and dynamic settings.

These activities are further split into two implementation cycles. The overall implementation approach initiated with the design of the data management PaaS framework, where the key interfacing requirements with the infrastructure's data sources at the lower end, and the end-user requests (through the data catalogue) at the higher end have been defined. The design has been redefined and updated once the overall AC³ platform architecture evolved and identified the ways that the data sources are used through the service deployment and management mechanism.

The second implementation cycle, that is about to begin (i.e. after M18), consists of the development of the software components that implement all the designed interfaces and integration mechanisms needed to realize the AC³ data management platform. In the next period of the project, all the Use Case data sources will be integrated with the AC³ infrastructure and will be introduced to the AC³ Use Case applications in order to generate the needed results. Additionally, all the implementations and concepts, previously designed for ease of use and value to application developers, will be tested.

1.2 Link with other project activities

This deliverable reflects mainly the design and development activities of T3.4 and some aspects of T3.1 implementations related to the data catalogue.

Initial feedback has been received from the architecture definition task in WP2, on the overall design and the flow of information, as presented in D2.1 (Sections 4.8, 5.2-5.4) in M08. Further details on the related technologies to support the targeted architecture have been presented in D2.3 (sections 5.2, 4.3.2) in M12.

The data management mechanism has been further evolved and adapted to the overall platform requirements and implementation features during the first implementation phase and the work presented in this deliverable. This work will also be reflected in the updated final architecture in D2.2, in M24. Furthermore, direct feedback is provided to WP5 T5.1 for the design of the overall system and the interfacing with the AC³ service deployment and management mechanisms.

1.3 Mapping AC³ Outputs

The purpose of this section is to map AC³ Grant Agreement commitments, both within the formal Deliverable and Task description, against the project's respective outputs and work performed.

Table 1: Adherence to AC³ GA Deliverable & Tasks Descriptions

AC³ GA Component Title	AC³ GA Component Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
D3.3 Report on data management for applications in CECC – Initial			
TASKS			
T3.4 Applications data management as a PaaS	Initial Data management mechanism	Section 4 Section 5 Section 2	Interfacing and flow of data between the continuum and external data sources. Utilization of integrated data sources to drive the deployment and ongoing management of services Main purpose and implemented functionalities, and Mapping of data management in the overall architecture
T3.1 Applications data management as a PaaS	Connectivity with data catalogue	Section 3 Section 2	Provides The components that form the AC³ Data Catalogue, their interactions, and interfaces Mapping of data catalogue in the overall architecture

1.4 Deliverable Overview and Report Structure

Deliverable D3.3 is structured as follows:

- Section 2 provides a mapping of the data management related functionalities with the overall architecture, highlighting the main development and innovation areas.
- Section 3 presents the implementation plans and development progress with respect to the data catalogue, focusing on the adopted mechanisms and the considered data types.
- Section 4 provides insights into the methods used to interface with data sources, the technologies involved, the software components that will be implemented and used and the planned improvements for the next period of the project.
- Section 5 focuses on the methodology used to include the actual data retrieval and processing components in the AC³ application's lifecycle, the technologies involved and the software components that will be implemented and used.
- Section 6 concludes on the initial data management activities and summarises the key findings and developments

2 Data management in the overall architecture

Data management within the Cloud-Edge-Client Continuum (CECC) architecture represents a foundational pillar, essential for the orchestration, optimization, and deployment of applications across the continuum. This chapter delves into the intricacies of data management within this sophisticated architecture, drawing upon the foundational principles outlined in Deliverable D2.1 "1st Release of the CECC framework and CECCM" and incorporating the latest advancements and refinements to be detailed in Deliverable D2.2 "Final release of the CECC framework and CECCM". The subsequent sections will explore the main concepts and targeted functionalities of data management, the mapping of these functionalities within the AC^3 architecture, and the targeted areas for development to enhance the system's data management capabilities.

2.1 Main concept and targeted functionalities

The overarching concept of data management in the AC^3 architecture is to facilitate seamless, efficient, and secure interactions between diverse data sources and the continuum's applications and services. This is achieved through a comprehensive framework that integrates advanced data handling, processing, and exposure functionalities with the application deployment and management processes. This innovative approach ensures that data is not only accessible but also actionable, within the overall service lifecycle and dynamically adapted to the end user application specific requirements, the application profile and the infrastructure status.

In this context, AC^3 introduces a significant innovation by integrating data management as a Platform as a Service (PaaS) within the CECCM. This initiative aims to simplify data handling within the cloud management system, making it easier for developers to work with data resources. By embedding essential components for data collection, processing, and management within a PaaS framework, the project aims to create a cohesive environment for managing computing, data, and network resources, fostering smoother integration and collaboration among stakeholders.

At the core of this initiative are the AC^3 Data Catalogue and the AC^3 data management Application Add-ons, which streamline the complexities associated with data access and handling. These components encompass a range of functionalities, including data indexing, retrieval, parsing, storing, and streaming from the data sources integrated with AC^3. They act as a bridge between application developers and diverse data sources, providing intuitive APIs while supporting various data formats and structures. Additionally, the AC^3 Data Catalogue includes features like automated data discovery mechanisms, facilitating the identification and categorization of data sources.

The AC^3 Data Management PaaS is designed to address a range of critical functionalities aimed at enhancing data management within the CECC ecosystem. More specifically, the platform is designed to ensure:

a) **Efficient data integration and interoperability**, seamlessly connecting heterogeneous data sources and sinks. This promotes a fluid data environment where information moves smoothly across different parts of the system.
b) **Dynamic data discovery and cataloguing features** to complement data integration and interoperability. These mechanisms provide users with intuitive and easy access to data resources, streamlining the process of finding and utilizing data.
c) **Advanced data processing and analysis**, enabling sophisticated capabilities that allow for real-time insights and informed decision-making. This is crucial in environments where speed and accuracy are paramount.
d) **Secure and controlled mechanisms for data exposure and sharing**, ensuring that all interactions with data meet stringent privacy and compliance standards. This is essential in maintaining trust and safeguarding sensitive information.

e) **Scalable and resilient infrastructure**, which is designed to adapt to evolving data volumes and processing demands, ensuring that it can handle growth and changes without compromising performance or stability. This scalability is vital for organizations that anticipate increases in data as they expand or as their operational needs evolve.

Together, these functionalities make the AC^3 system a comprehensive and robust tool for managing data across diverse and dynamic environments.

2.2 Mapping to the AC^3 architecture

In the AC^3 architecture, the data management features are closely integrated into the platform's architecture design, ensuring that data can be effectively captured, processed, and utilized across various layers and components. Key elements and components of this architecture include:

(a) the **Data Catalogue** that provides a comprehensive inventory of available data, including information about formats, types (live or stored), locations, and access methods. It offers all the necessary procedures and interfaces to register, federate, manage, and provide access to data spaces of application developers or third-party IoT infrastructure providers.

(b) **Data Sources (Hot and Cold)** refer to the locations where the actual data available to the applications of AC^3 reside. These are categorized based on their access frequency, usage patterns, and lifecycle. **HOT** data is frequently accessed and requires fast retrieval times; often stored in high-performance storage for immediate use, with a live feed nature, where each new measurement is added to the available dataset once produced. Time is crucial for this type of data, and it needs to be processed as fast as possible. **COLD** data is accessed less frequently, stored in cost-efficient storage solutions, and may require more time to retrieve when needed based on its size and format.

(c) **Data Connectors (for Data Sources)** serve as the vital links that facilitate the flow of data between the continuum and external data sources. These connectors are engineered to accommodate a diverse array of data types, formats, and communication protocols, ensuring that data can be ingested from and dispatched to a wide range of sources, including IoT devices, cloud repositories, and edge computing platforms.

(d) **Data Management Application Add-ons** are components that live and operate inside the application's premises with a set of roles to access, retrieve, and process the data stored in the available **Data Sources**. These components include the following:

- **Data Connectors (for Data Consumers):** Aimed at delivering data to end-users or downstream systems, these connectors ensure that data consumers can access and utilize the data efficiently. They adapt data delivery to meet the specific needs and consumption patterns of different consumers, optimizing user experience and data utility.
- **Data Mappers**: They play a crucial role in ensuring the seamless flow and integration of data across various components of the architecture. By transforming data formats and structures, Data Mappers enable different systems to understand and utilize data effectively, facilitating interoperability and reducing incompatibility issues.
- **Message Broker**: Serves as a communication channel within the CECCM, facilitating data flow and exchange between different components. It is a central point that receives data requested by applications executed in the CECCM and delivers them to the appropriate applications.
- **Data Manipulator** This component is responsible for refining and preparing data for further use. It performs essential operations like cleaning, transforming, and aggregating data, making it more valuable and suitable for analysis, decision-making, or other specific purposes within the architecture.

An updated version of the AC³ architecture (D2.1) is presented in Figure 1, where the data management framework elements are included and highlighted.

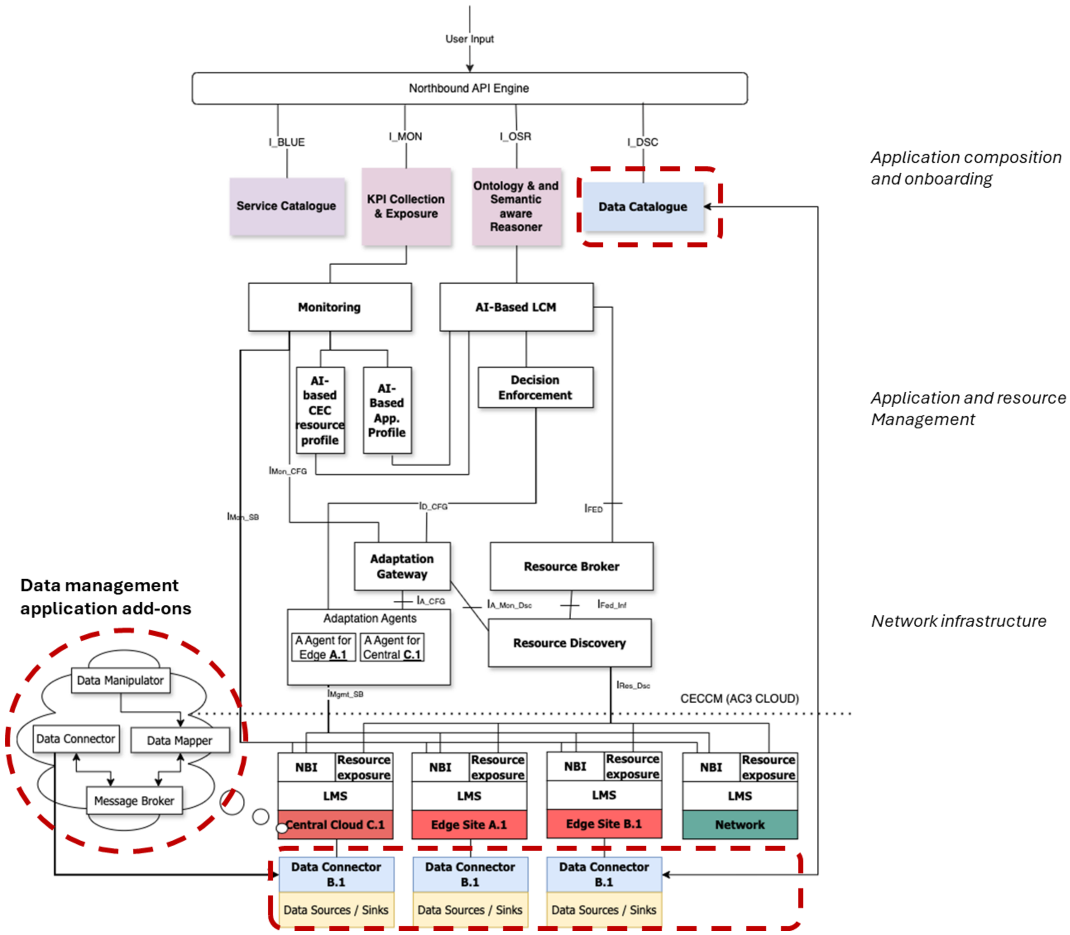

Figure 1 The updated AC3 Architecture with the data management framework elements being highlighted

The data sources are related to the interconnected infrastructures, whether these reside at the edge/far-edge, core, or cloud part of the network. The exposure of the data resources, type, and data information is done during the infrastructure registration process and is considered known to the AC³ framework when service instantiation takes place. This means that the data catalogue residing in the upper application composition and onboarding layer contains the targeted infrastructure data information required during the composition of the service request (i.e., service descriptor model).

The data sources within the registered infrastructures provide their data through specific data connectors. Such connectors include and may not limited to: a) provider-specific cloud and storage connectors; b) IoT device connectors such as MQTT, CoAP, lightweight M2M, Modbus, etc.; c) Edge computing connectors; d) database connectors. The interconnectivity with any data connector is subject to the related application component or an intermediate interfacing layer that may be adopted by a certain infrastructure. In AC³, the focus is on the support of specific types as defined by the supported use case examples; however, in principle, the framework can be adapted to any connector type, provided that the appropriate controller is made available.

Finally, the data management application add-ons are not explicitly related to the AC³ architecture as separate architectural modules. They are application-related data usage functions and are considered part of the deployed and dynamically managed end-user application. This allows certain data mapping and manipulator functions to be implemented separately per application while also each application handles its potential connectivity with other external data sources. The scheme also enables data-related monitoring information to be exposable to the application management layer for dynamic data-related configuration.

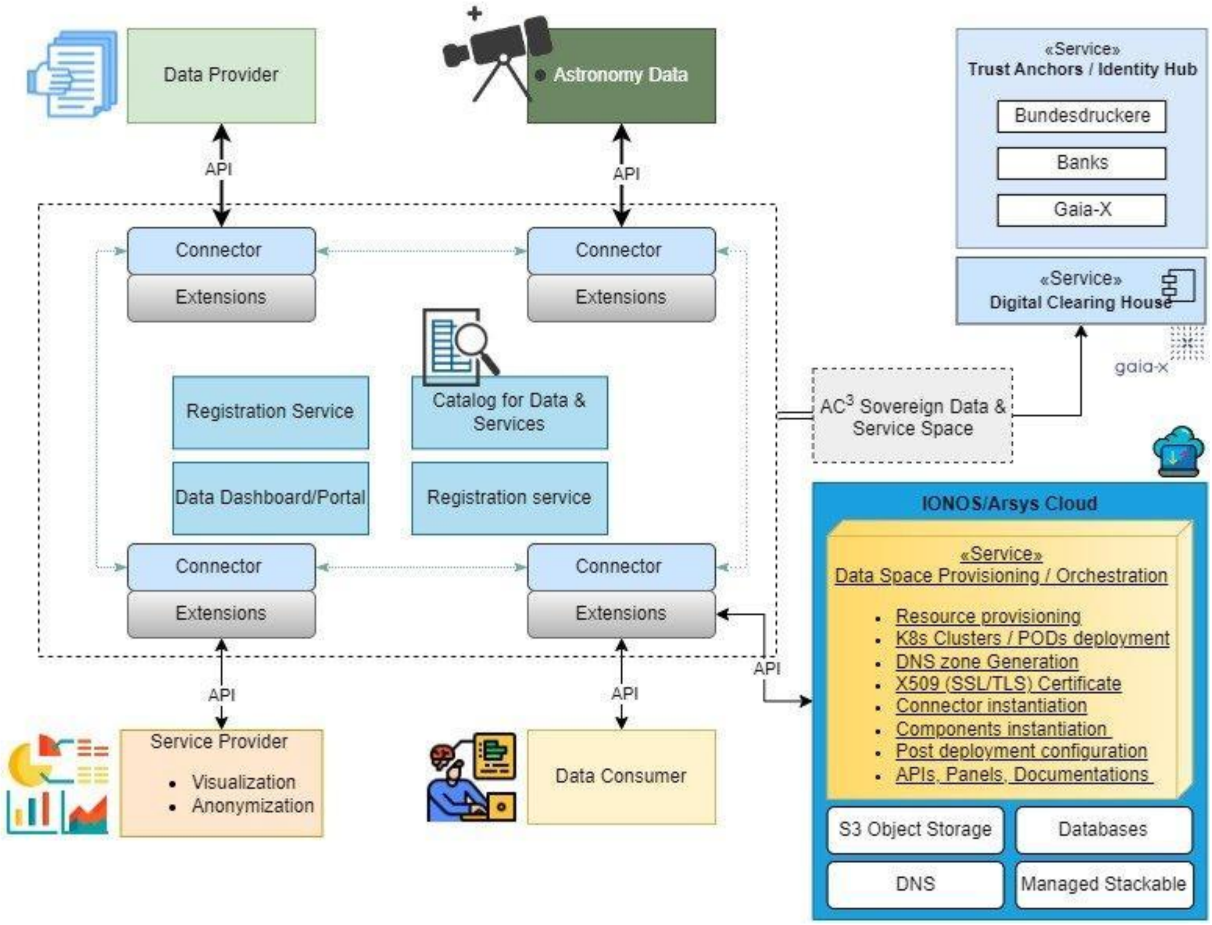

Figure 2: Conceptual diagram of AC³ Data Space

In order to further highlight the interconnection of the AC³ CECC architectural framework with the dataspace ecosystem, a sophisticated dataspace concept is illustrated in Figure 2, comprising various essential components and their interactions. At the core of the diagram are data providers represented by "Data Provider" and "Astronomy Data" boxes (as an example for the 3rd Use Case of AC³) serving as sources supplying data through APIs (more precisely, the Data Source Connector components to be presented in the next sections). The data space connectors (i.e., in AC³ consortium adopted EDC connector), depicted by four boxes with extensions, play a pivotal role in interfacing with both data providers and consumers to ensure secure and managed data exchange within the system. These data space connectors act as bridges facilitating seamless communication and data flow between various entities in the dataspace.

Furthermore, the architecture includes key components such as the Registration Service and Data Dashboard/Portal, which can serve as central services for managing and registering data sources. These components not only facilitate user access but also can provide a centralized dashboard for data insights and management. Additionally, the Catalogue (i.e., Piveau[1] catalogue) for Data & Services acts as a vital repository,

[1] https://www.piveau.eu/en/

collaborating with the registration service to offer a comprehensive overview of available data. Service providers are instrumental in offering additional services like visualization and anonymization, leveraging the data from data space connectors and registration services to enhance data before reaching the end data consumers who interact with the system through APIs.

The data flow within this framework is meticulously structured to ensure efficient data ingestion, management, and utilization while focusing on identity and trust services for secure and compliant data exchanges. External services like Trust Anchors/Identity Hub and the e-Services Digital Clearing House play a crucial role in managing identity and trust services. The Sovereign Data & Service Space, governed by AC^3 and supported by IONOS-Arsys Cloud infrastructure and the Spark Works IoT Analytics Engine, provides a managed and reliable environment for data provisioning as well as orchestration and processing. Finally, the IONOS-Arsys Cloud offers a range of cloud services essential for the infrastructure, including resource provisioning, deployment, DNS management, SSL/TLS certificates, S3 object storage, and other critical services required for the seamless functioning of the dataspace ecosystem and securely share the data between different stakeholders of AC^3 in a sovereign, manner. Overall, this comprehensive framework underscores the importance of secure, sovereign, managed, and efficient data exchange among diverse stakeholders in a structured and robust dataspace concept.

2.3 Targeted development areas

The continuous evolution of the AC^3 architecture's data management capabilities is focused on three key development areas: interfacing with end users, interfacing with data sources, and use of data sources.

2.3.1 Interfacing with End Users

This area focuses on the mechanisms through which end users interact with the data managed within the AC^3 architecture. The data catalogue provides a centralized platform for users to search, discover, and access available data assets. Simultaneously, data exposure functionalities ensure that users can securely and efficiently utilize these data assets within their applications and services, in compliance with relevant data governance and privacy policies. The goal in this domain is to enhance the user interface and experience of interacting with the data catalogue to make data discovery and access more intuitive and efficient, strengthen data exposure mechanisms to support secure, compliant, and flexible data sharing, enabling users to leverage data assets across various applications and services.

2.3.2 Interfacing with Data Sources

In the context of the AC^3 architecture, Data Connectors serve as vital links that facilitate the flow of data between the continuum and external data sources. These connectors are engineered to accommodate a diverse array of data types, formats, and communication protocols, ensuring that data can be ingested from and dispatched to a wide range of sources, including IoT devices, cloud repositories, and edge computing platforms. The primary goal of the development efforts is to enhance the adaptability and performance of data connectors, ensuring they can efficiently manage increasing data variety and volume. This involves incorporating advanced data handling technologies, improving real-time data streaming capabilities, and ensuring robust error handling and data integrity checks. In the context of AC^3, we will focus on providing connectors for streaming IoT data, streaming video contents, and previously stored video and astronomical observation files. These data types represent a wide range of data sources that can provide us with a great understanding of what different data types need in terms of implementation and a great pathway on how to provide additional implementations for other users in the future.

2.3.3 Use of Data Sources through Service Deployment and Management

This development area emphasizes the utilization of integrated data sources to drive the deployment and ongoing management of services within the AC^3 architecture. By leveraging real-time and historical data insights, services can be dynamically adapted to meet changing user needs, optimize resource usage, and enhance overall service performance. The main target is to develop capabilities for dynamic service deployment that utilize integrated data insights for real-time adaptation and optimization. Enhance service management frameworks to incorporate data-driven decision-making processes, enabling continuous service improvement and ensuring optimal alignment with user requirements and system constraints.

3 Data sharing and usage by end users

This section presents the implementation plans and development progress with respect to data sharing and usage as this is primarily achieved through the data catalogue. The focus is on the adopted mechanisms and, in particular, the considered data types and the data catalogue interfacing.

3.1 Overview of the implementation plan

Sharing data in the context of the AC³ Data Management Platform as a service is achieved using the AC³ Data Catalogue service. This service is instrumental in providing all the appropriate descriptions, metadata, information, and guidance to software developers regarding the use of the offered datasets and data sources. These data sources, as mentioned before, incorporate a broad spectrum of domains and need to be properly described and annotated to be easily searched and discovered. The annotation of the data sources is done using rich semantic information based on well-established ontologies provided by the owners of the data during the initial phase of the data source registration. An example screenshot of the AC³ Data Catalogue is available in Figure 3. In this portal, users can browse and check the datasets shared, initially, by the Use Cases of AC³ and, later on, any other datasets made available through our platform.

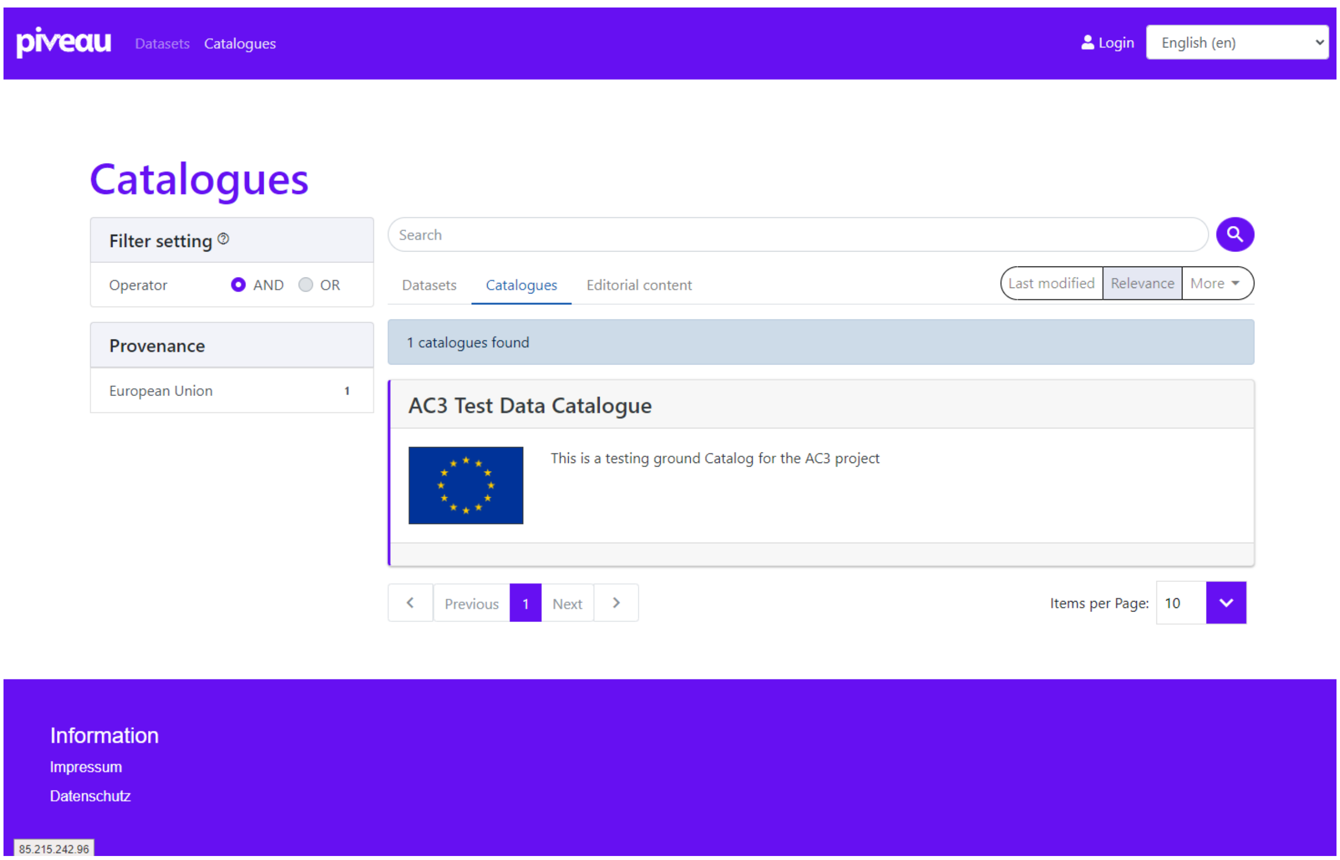

Figure 3 The AC3 Data Catalogue

The adopted approach in AC³ is based on the adoption of a well-established and widely used solution like Piveau, instead of re-inventing the wheel and providing yet another incomplete solution for our users. Piveau is a European data management ecosystem that provides components and tools to support the entire data processing chain from harvesting, aggregation, provision, and use, focusing on open standards and high interoperability. Our goal is to adapt Piveau to AC³ needs and provide these changes as possible extensions to the Piveau maintainers as an additional contribution of AC³ to the European Open-Source software community increasing the value, reach and potential of these tools.

The AC³ Data Catalogue therefore is composed of the following internal components:

- The Piveau Hub UI, a component that acts as a frontend service through which users can interact with catalogues and datasets.
- The Piveau Hub Repository, the core module for the Piveau data platform that manages and syncs the triple store and the search index, providing a rich RESTful API for interacting with the catalogue's items.
- The Piveau Hub Search, an interface on top of the RDF Triple Store and the Elasticsearch instances used for indexing on the stored datasets and catalogues.
- A Virtuoso RDF Triple Store, for storing all the semantic information of the system.
- An Elasticsearch instance, which is used for indexing the stored data and providing advanced search capabilities.

On top of that, in order to allow users of AC³ to interact with the data catalogue of the project, we use a Keycloak Server for implementing all the authentication, authorization and accounting in the Data Management PaaS. Keycloak is also a well-established Open Source Identity and Access Management software that has already been integrated with Piveau.

Figure 4 showcases the interactions between the components that formulate the AC³ Data Catalogue based on the above description and the interfaces they expose.

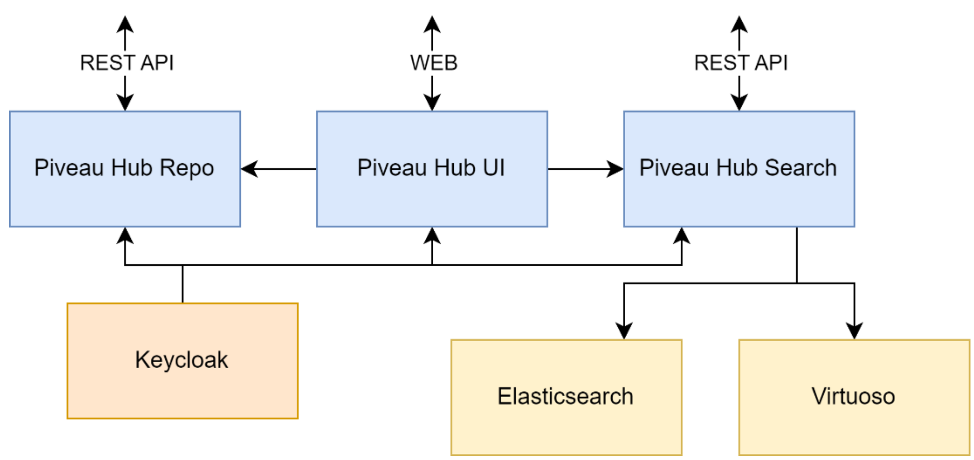

Figure 4: The components that form the AC³ Data Catalogue, their interactions, and interfaces

3.2 Initial developments

The initial phase focused on obtaining a thorough understanding of the Piveau Hub data catalog and identifying the suitable data attributes required to effectively represent the three AC³ use cases within the system. This process began with an extensive examination of the metadata associated with the Piveau Hub data catalogue, encompassing data formats, types, and schemas, ensuring alignment with the AC³ use cases. Subsequently, a careful assessment was conducted on the essential data attributes for each use case, taking into consideration factors such as volume, complexity, and compatibility with the proposed system architecture. Consequently, the AC³ data catalogue was structured based on these identified data sources, while tailored datasets were prepared for each unique use case to facilitate efficient access and utilization within the system.

3.2.1 Data Types of project Use Cases

Regarding the data to be stored and made available through the AC^{3} Data Catalogue, we have the following:

- IoT device data collected through the IoT Infrastructure of Use Case 1. This data is Hot data collected from the sensor deployments in real-time and made available as data streams to application developers. Cold datasets based on the collected data will also be made available for testing.
- Video feed data that are collected through the video cameras deployed for Use Case 2. This data will be offered as Hot and Cold data to application developers.
- IoT device data collected through the IoT Infrastructure of Use Case 2. This data will be offered as Hot and Cold data to application developers.
- Astronomical observations are made available as Cold datasets in the context of Use Case 3.

Table 2 Description of the Data Types to be handled in the context of AC^{3} based on the Use Cases of the project

Data Type	Description	UC1	UC2	UC3
Time-Series	Represents measurements or observations recorded over time, allowing for analysis of trends, patterns, and anomalies.	HOT COLD	HOT COLD	
Video Stream	Represent a live or post-processed continuous flow of audio and video from a deployed camera.		HOT COLD	
DataCubes	Three-dimensional array of data values, incorporating two spatial axes that represent a specific region of the celestial sphere			COLD

3.2.2 Data catalogues and datasets in the AC^{3} Data Catalogue

As part of our initial work on representing the datasets of our 3 use cases on the AC^{3} Data Catalogue we have defined the basic parts needed to represent data as well as datasets from their data sources. In this subsection, we present the initial descriptions of these components that will be used in the future. Datasets and data sources in Piveau are represented using **Resource Description Framework (RDF)** descriptions. More specifically, a textual syntax for RDF called **Turtle** that allows an RDF graph to be completely written in a compact and natural text form, with abbreviations for common usage patterns and datatypes. Turtle provides levels of compatibility with the **N-Triples**[2] format as well as the triple pattern syntax of the **SPARQL W3C Recommendation**[3].

Each document is split into two parts. The first part, contains all **@prefix** lines that define prefixes that make it easier to reference **Internationalized Resource Identifiers (IRIs)** in the serialization of the graph. The second part of the document is the actual triples that describe the RDF document with information on the dataset and data source they describe.

Listing 1 Dataset Semantic description in Turtle RDF format

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix locn: <http://www.w3.org/ns/locn#> .
```

[2] https://www.w3.org/TR/turtle/#bib-N-TRIPLES
[3] https://www.w3.org/TR/sparql11-query/

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix time: <http://www.w3.org/2006/time> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .


<https://piveau.io/id/catalogue/test-catalog>
   a dcat:Catalog ;
   dct:type "dcat-ap";
   dct:title "AC3 Test Data Catalogue"@en ;
   dct:description "This is a testing ground Catalog for the AC3 project"@en ;
  dct:language  <http://publications.europa.eu/resource/authority/language/ENG> ;
   dct:spatial <http://publications.europa.eu/resource/authority/country/EUR> ;
   dcat:dataset <https://piveau.io/set/data/test-dataset> ;
   dcat:record <https://piveau.io/set/record/test-dataset> ;
   dct:publisher  [
      a foaf:Agent ;
      foaf:homepage  <https://ac3-project.eu/> ;
      foaf:mbox      <mailto:info@ac3-project.eu> ;
      foaf:name      "AC3 Project"
   ].
```

Similarly, we work on providing templated descriptions for the datasets that are to be used for each of the use cases of AC^3. These descriptions will need to be improved and enriched as the project progresses in order to provide more rich and accurate information to be used in conjunction with the rest of the AC^3 platform. An example of this description is presented in Listing 2. In this snippet, we can see multiple fields regarding the data available, their meta information, the location the data refer to, their supplier, the period of data collection, the license provided, as well as the means of distribution the application developer needs to follow in order to get access to the data.

Listing 2 Dataset Semantic description in Turtle RDF format

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <http://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix time: <http://www.w3.org/2006/time> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix edp: <https://europeandataportal.eu/voc#>


<https://example.com/id1>
   a dcat:Dataset ;
   dct:title "Example Dataset 1"@en ;
   dct:language  <http://publications.europa.eu/resource/authority/language/ENG> ;
   dct:description "This is an example Dataset" ;
   dcat:theme <http://publications.europa.eu/resource/authority/data-theme/ENVI> ;
   dct:issued "2024-04-24T00:00:00"^^xsd:dateTime ;
   dct:modified "2024-04-24T00:00:00.000000"^^xsd:dateTime ;
   dct:accessRights ":public" ;
   dct:spatial [
      a dct:Location ;
      locn:geometry "<gml:Point srsName=\"http://www.opengis.net/def/EPSG/0/4326\"><gml:Pos>53.1485
12.915</gml:Pos></gml:Point>"^^<http://www.opengis.net/ont/geosparql#gmlLiteral>
   ] ;
   dct:temporal [
      a dct:PeriodOfTime ;
      schema:endDate "2015-06-09T00:00:00"^^xsd:dateTime ;
      schema:startDate "2015-06-09T00:00:00"^^xsd:dateTime
   ] ;
   dcat:contactPoint  [
      a vcard:Kind ;
      vcard:hasEmail  <mailto:d.amaxilatis@sparkworks.net> ;
      vcard:fn   "Dimitrios Amaxilatis"
   ] ;
   dct:publisher <http://publications.europa.eu/resource/authority/corporate-body/COUN_COOP_EU_SYR> ;
   dcat:keyword
      "piveau", "test", "open data", "ac3" ;

   dct:conformsTo "Conforms To String" ;
   dct:conformsTo [
      rdfs:label "Conforms To"
   ] ;
   dct:conformsTo <http://conforms.to/conformation> ;
   dct:conformsTo <https://example.de/def/conformsto> ;
   dct:provenance [
      a dct:ProvenanceStatement ;
      rdfs:label  "This is a provenance statement"@en
   ] ;
   foaf:page <http://www.documentation.com> ;

   dcat:distribution <http://example.com/distribution/1> ;
   dcat:distribution [
      a dcat:Distribution ;
```

```
    dcat:accessURL <http://sparkworks.net/data.csv> ;
    dct:format <https://example.de/def/format/CSV> ;
    dct:license [
      dc:identifier "test-licence" ;
      skos:altLabel "Test Licence" ;
      skos:prefLabel "This is a Open Test Licence" ;
      skos:exactMatch <http://testlicence.org>
    ];
    dcat:mediaType <http://www.iana.org/assignments/media-types/text/csv> ;
    dct:title "Example Distribution 1" ;
    dct:description "This is a example Distribution 1"
  ] .
```

3.3 Planned implementations and extensions for the final phase

3.3.1 Linking Data Catalogue with the AC³ Service Catalogue

Parallel to the Data Catalogue, the Service Catalogue lists and describes the various services available within the CECC environment. It acts as a resource for users to explore available applications and services, understand their functionalities, and initiate their deployment. The Service Catalogue plays a significant role in maintaining a clear inventory of the ecosystem's capabilities, aiding in the seamless integration and management of services.

The Data Catalogue's detailed inventory of data assets is aligned with the Service Catalogue's listings of applications and services using common semantic ontologies to describe the data types needed. This alignment ensures that every service listed in the Service Catalogue can be associated with relevant data sources defined in the Data Catalogue. This linkage facilitates the seamless deployment of services by automatically associating them with the necessary data inputs as defined in the Service Catalogue. It allows service managers to understand the data dependencies of each service, streamline configuration processes, and ensure that all data requirements are met before service deployment.

3.3.2 Linking Data Catalogue to Semantic Aware Reasoner

The Ontology Semantic Reasoner (OSR) operates as an advanced reasoning engine that interprets and processes the information structured in the Data and Service Catalogues using semantic web technologies. It applies logic and reasoning to understand and adapt the interactions between various system components and data layers. The OSR translates high-level policy and intent descriptions into actionable configurations, enhancing decision-making processes and application management across the CECC.

The integration of the Data Catalogue with the Semantic Aware Reasoner (OSR) is a multi-step process involving the linkage of the Data Catalogue to the Service Catalogue, which is then utilized by the OSR. This integration is important for aligning data access and usage with the strategic goals of the CECC environment, ensuring that data-driven decisions and operations are both efficient and aligned with overarching policy frameworks.

Once the Data Catalogue is intricately linked to the Service Catalogue, the fused repository transforms into a knowledge base for the OSR. This integrated catalogue serves not just as a data store, but as a dynamic and intelligent tool that the OSR uses to understand and process relationships and dependencies between various services and their corresponding data sources. The OSR analyses the structured relationships defined by the ontologies in the combined catalogues. These relationships outline how different services interact with various data types, their sources, and the required data formats and access protocols. Additionally, it interprets policies and user intents to manage how services should interact with data. This involves parsing high-level policy documents and user requirements to derive actionable configurations and operational rules.

3.3.3 Finalize the data sources and datasets of the project

One of the pivotal tasks for the forthcoming period is to finalize the datasets each of the AC^3 Use Cases will offer. This endeavour is vital as it will not only display the diverse functionalities of the AC^3 Data Catalogue but also clarify the resources available to application developers. We must meticulously define the specific datasets featured in each Use Case, along with their distinct characteristics. For instance, within UC2, video feeds might be provided either as live streams or as pre-curated video files, which are particularly suitable for application training scenarios.

Each dataset's precise number, duration, and volume will be determined through close collaboration with the use case owners. This collaborative approach ensures that the datasets are tailored to meet the strategic objectives and operational needs of each Use Case, enhancing their practicality and relevance.

Additionally, an integral part of this task includes the careful selection of appropriate ontologies that best describe the data types being offered, as well as the sources of these data. For example, UC3 will require ontologies that cover astronomical observations to accurately reflect the complexity and specificity of space data, whereas UC2 will utilize ontologies related to video data processing, and UC1 will focus on ontologies appropriate for Internet of Things (IoT) data.

Selecting the right ontologies is crucial because they provide a structured vocabulary that enhances data interoperability and understanding across different systems and stakeholders. By aligning each Use Case with the most suitable ontologies, we can ensure that the AC^3 Data Catalogue is not only robust and comprehensive but also adaptable to the evolving needs of developers and researchers in various fields.

3.3.4 Align the ontologies used in the Data Catalogue with the Service Catalogue

Another important task is to align the ontologies described above with the ontologies used by other important AC^3 components, like the Service Catalogue, the Ontology Semantic Reasoner (OSR) and the Life Cycle Management (LCM). The description provided for each dataset, will be used by the OSR to select the appropriate datasets to be used, and the LCM to forward them to the deployed applications to use them during their lifecycle, while the Service Catalogue will use these ontologies to visualize the datasets and facilitate searching them by the application developers.

Aligning semantic ontologies between two software components involves several key steps to ensure seamless data exchange and accurate interpretation. Initially, both ontologies will be reviewed and documented to understand existing terms, relationships, and structures. A gap analysis will identify discrepancies and overlaps, followed by the development of mappings to align differing terms and structures. These mappings will be integrated, and comprehensive testing will ensure the integrity and consistency of data exchange, while documentation of the aligned ontology and standardization according to industry norms will be essential for clarity and broader application. Ongoing maintenance procedures will be established to update the alignment based on changes in component ontologies. The successful completion of this task enhances interoperability, reduces semantic barriers, and improves overall system performance, facilitating smoother data interactions across integrated software components.

4 Interfacing with data sources and data collection mechanisms

The first task for enabling the processing of data in the context of AC³ is the integration of data sources with the AC³ infrastructure. This section provides insights into the methods used to interface with data sources, the technologies involved, the software components that will be implemented and used, and the planned improvements for the next period of the project.

4.1 Overview of the implementation plan

4.1.1 Adopted technologies

The main component that takes the role of providing access to the integrated data sources is called a Data Connector. The Data Connector is not a single piece of software but a concept that facilitates all the interactions needed to register the data available to the AC³ Data Catalogue and to serve the available data to the applications of AC³.

4.2 Initial developments

4.2.1 Data source registration & maintenance

The registration of data sources needs to start from the AC³ Data Catalogue, as presented in the previous section. This basic information provides the data source with a dedicated identifier for each dataset that is made available to application developers. This identifier is used then by the Data Source Connector to provide updates to the dataset's description through the Piveau Hub Repo API. The API simplifies the management of datasets, adhering to the DCAT-AP standard for structured data descriptions. It integrates seamlessly with Virtuoso, an efficient database server, and Triplestore, making dataset management more accessible. The DCAT-AP standard ensures consistent, comprehensive data descriptions, enhancing data sharing and discovery. The API simplifies the process of making datasets accessible to a wider audience. This promotes transparency, knowledge sharing, and collaboration. This introduction lays the groundwork for exploring the Piveau Hub Repo Open API. Detailed information on authentication, API endpoints, error handling, and practical examples can be found in the documentation of the Piveau Hub Repo[4].

4.2.2 Data Source Connectors

The next step in our work is the implementation of the Data Source Connectors. These components will be based on the well-established Eclipse Data Space Connector (EDC)[5], which is designed as an open-source framework aimed to facilitate the controlled and secure sharing of data in decentralized environments. Especially relevant for projects like the AC³ consortium, the EDC provides a suite of functionalities that support the handling and distribution of IoT and video data—distinguishing between different data temperature states like hot and cold and addressing the varying processing needs for these data types. Amongst others, the EDC offers the following functionalities:

- **Real-time Data Streaming:**
    - **IoT and Video Integration:** EDC supports the integration and management of streaming data from IoT devices and video feeds, enabling real-time data ingestion and processing.
    - **Efficient Data Routing:** Data streams are dynamically routed based on policies, client requirements, and data types, optimizing for latency and bandwidth.
- **Hot and Cold Data Management:**

[4] https://piveau.ac3-project.eu/
[5] https://projects.eclipse.org/projects/technology.edc

  - **Hot Data Handling:** For 'hot' data, which requires immediate analysis for time-sensitive decisions, EDC facilitates fast access and low-latency processing.
  - **Cold Data Storage:** 'Cold' data, not frequently accessed, is managed differently, optimizing storage techniques and cost efficiency and using more cost-effective, scalable storage solutions.
- **Secure Data Sharing:**
  - **Governance and Compliance:** The framework ensures that data sharing adheres to specified governance policies and compliance requirements, crucial for sensitive IoT and video data.
  - **Data Access Controls:** EDC incorporates sophisticated user and data access controls, encryption, and secure data transmission protocols to protect the integrity and confidentiality of data as it moves across different jurisdictions or within various sectors of the industry.

As part of the AC^3 project, we aim to leverage EDC's capabilities to build a robust data space on top of the IONOS and Arsys cloud infrastructure. The IONOS-Arsys infrastructure offers reliability, scalability, and performance, which are essential for effectively implementing a data space that handles large-scale, high-speed data streams and facilitates complex transactions and interactions between diverse participants. Additionally, IONOS and Arsys provide scalable cloud infrastructure that can dynamically support the intensive demands of streaming, processing, and storing both hot and cold IoT and video data. The deployment of our solution on the IONOS-Arsys infrastructure enhances operational efficiencies, reduces latency, and improves data throughput, essential for the real-time data needs of the AC^3 Consortium.

In summary, the use of EDC in conjunction with the Piveau Hub within the AC^3 consortium, supported by IONOS-Arsys Cloud infrastructure, is poised to create an advanced, secure, and efficient data space that will enable effective data sharing and utilization, drive innovation, and foster collaboration across diverse sectors and stakeholders.

As a result, based on the use of the EDC we expect to provide the AC^3 end user and infrastructure owners with the following three connectors designed as extensions to the EDC: Streaming IoT Data Connector, Streaming Video Data Connector, and Cold Data Connector. These connectors will be further discussed and presented in the rest of this section.

4.2.2.1 Cold Data Connector

The **Cold Data Connector**, based on the *IONOS S3 EDC extension*, is a pivotal component designed to integrate seamlessly with the Eclipse Data Space Connector (EDC), specifically leveraging the robust and scalable storage capabilities of IONOS S3 services. Given the AC^3 project's focus on developing a secure and efficient data space, this extension plays a critical role by enhancing the functionality of the minimum viable data space setup. Let's delve into how the IONOS S3 EDC extension ensures the security and data integrity essential for the AC^3 project.

- **Key Functionalities**

Data Storage and Management: Leveraging IONOS S3's storage solution, the extension allows for efficient storage management practices such as data replication and lifecycle policies, which help manage data through its various stages effectively.

- **Scalability:** With IONOS S3, the EDC can scale storage resources up or down based on real-time demand without compromising performance or data availability, which is crucial for maintaining the robustness of data operations within the AC^3 project.
- **Integration Simplicity:** The extension is designed to integrate smoothly with minimal configuration, facilitating ease of setup and maintenance. This integration simplicity ensures that the focus remains on data utilization rather than infrastructure management.

- **Scalability:** With IONOS S3, the EDC can scale storage resources up or down based on real-time demand without compromising performance or data availability, which is crucial for maintaining the robustness of data operations within the AC^3 project.

- **Ensuring Security and Data Integrity**

The integration of the IONOS S3 EDC extension with the Eclipse Data Space Connector enhances security and data integrity in several significant ways:

- **Encryption:** Data stored using the IONOS S3 service is encrypted at rest, protecting sensitive data from unauthorized access. This aligns with the stringent security requirements of the AC^3 project, where data confidentiality is paramount.
- **Access Control:** IONOS S3 services provide sophisticated access control mechanisms that can be configured to comply with the data governance policies of the AC^3 project. This includes setting permissions for data creation, modification, and deletion, which are crucial for maintaining data integrity and adherence to regulatory standards.
- **Audit Trails:** The ability to monitor and log access and activities related to the data stored in IONOS S3 helps in creating a transparent and traceable data handling environment. This is pivotal for audits and compliance checks, ensuring that all data transactions within the space maintain a high standard of accountability.
- **Data Resilience**: IONOS S3 offers high durability of stored objects, safeguarding against data loss due to system failures. This resilience is crucial for building a reliable data space where continuity and availability of data are maintained.

4.2.2.2 Streaming IoT Data Connector

The **Streaming IoT Data Connector** covers a large type of data sources to be integrated in AC^3, including various time-series-based data sources. It is a pivotal component designed to ease the introduction of real-time data for applications that need to process information generated from various IoT and sensing infrastructures, like building monitoring, utility metering, power grids, Industrial IoT deployments and others. Some of the key functionalities of this Connector will provide the ability to distribute the collected IoT data in real-time to client applications of AC3, providing scalability, ease of integration, and data security.

The implementation of the Streaming IoT Data Connector is planned for the next phase of the project.

4.2.2.3 Streaming Video Data Connector

The **Streaming Video Data Connector** covers specific data source connectivity needs in which live video data is collected from deployed on-site cameras, aggregated in the designated data space, and shared using the AC^3 infrastructure through the deployed application components. The 2nd use case of AC^3 is specifically designed to include this option. The connector design follows a similar approach to the IoT Data Connector with tighter latency and bandwidth requirements to ensure that the video feeds are delivered as fast as needed to their consumer applications. Some of the key functionalities this Connector will provide includes the ability to distribute the collected video feed data in real time to client applications of AC^3 providing scalability, ease of integration, data security, and user privacy.

The implementation of the Streaming Video Data Connector is planned for the next phase of the project.

4.2.2.4 Impact on the Minimum Viable Data Space

The minimum viable data space model aims to establish a foundational data space with essential capabilities operational. Here, the AC^3 EDC extensions enrich this model by providing a reliable, scalable, and secure data storage solution that aligns with the critical objectives of the project. By ensuring that the data managed within the space is both secure and robustly stored, the extension provides a critical backbone supporting the scalability and integrity of the project's data management practices.

In conclusion, the AC³ EDC extensions are not merely a tool for data storage and transfer; they are a strategic enhancement that bolsters the security, scalability, and efficiency of the AC³ project's data space. This setup provides the necessary groundwork for extending the capabilities of the data space, facilitating broader participation and more complex data interactions in future scales.

4.3 Planned implementations and extensions for final phase

The implementation and evaluation of the functionality and performance of the Data Connectors is the main planned implementation in this part. We will focus on developing the EDC extensions for the Streaming IoT Data and Streaming Video Connectors as well as expanding the implementation of the Cold Data Connector based on the IONOS S3 EDC extension.

5 Data management through service deployment and orchestration

The final step toward building data processing applications in the context of AC³ is to include the actual data retrieval and processing components in the AC³ application's lifecycle. This includes both the orchestration of the application, and its deployment and migration from Cloud to Edge and back, as needed. This section provides insights into the methodology used to achieve this, the technologies involved, the software components that will be implemented and used, and the planned improvements for the next period of the project.

5.1 Overview of implementation plan

Data processing in AC³ is implemented as part of the deployed applications. For this purpose, we provide application developers with specific application addons and guidelines for developing their own application addons that can connect to the available data sources, consume the data available and generate new data. The data processing application addons are expected in the form of deployable micro-services that provide target specific extensions to applications. To this direction, the existence of 3 very different use cases allows us to investigate different requirements and data settings that will help us designing our data processing components in a way that is best fitted to as may data sources and application needs. As all AC³ applications are based on a micro-service architecture, our addons operate following the same logic, providing multiple microservices that can be merged to provide application developers with all the building blocks they would need.

For the design of data processing micro-services we have defined two main approaches: the *Hot data approach* for live data sources, and the *Cold data approach* for datasets stored before usage.

5.1.1 Micro-service-based data management application design

The general approach is showcased in Figure 5 and represents the generic design of the data processing application addons. As part of the AC³ application, the AC³ CECCM can deploy the following components to allow for the processing of data in the context of the application: The Data Connector, a Message Broker, a Data Mapper, a shared storage location and one or more Data Manipulators. The role of each addon will be presented for each setting related to the data approach (Hot or Cold), in the rest of this section.

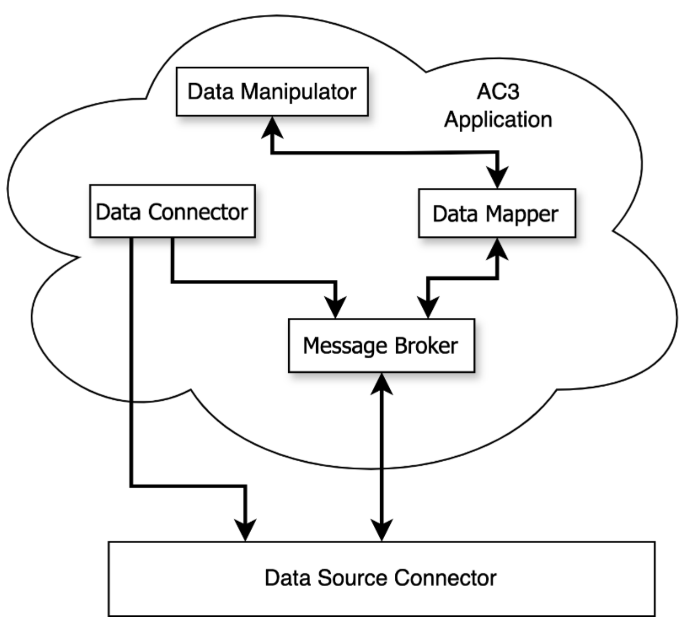

Figure 5: High Level View of an AC3 Data processing application

5.1.2 Adopted technologies

A core technology that is employed to allow for the transfer of data from data sources to data consumers in AC^3 are the Data Connectors. These software components, as described in the previous section, assume the responsibility of interfacing the Data Connectors on the Data Source's side and controlling the data retrieval for both the Hot and Cold data available.

Additionally, tools for streaming and transferring the collected data are defined and developed. Message Brokers like RabbitMQ[6] and Kafka[7], based on the Advanced Message Queuing Protocol (AMQP) and other protocols are used to allow for secure, asynchronous, and reliable data transfer, regardless of the actual size of the data in question.

Finally, due to the wide variety of the applications that need to be deployed (visible through the AC^3 use case needs, but also considering any generic deployment scenario), specific end-user defined application components will be used to do the actual data processing. However a common design feature in all these application components should be their ability to consume the retrieved data source data, as well as the need to provide the necessary feedback to the AC^3 platform for its operation. The implementation of the AC^3 data management addons is targeting to offer this common pool of components to service developers (or potentially providers) in order to adapt their processing application components to the data management application addons and successfully incorporate them in their design.

5.2 Initial developments

As mentioned above, the data management application addons are tailored to the two main types of managed data (i.e. hot and cold data), and therefore are presented separately for each case.

5.2.1 Hot Data management application addons

In a Hot Data setting, the AC^3 application employs a structured approach as illustrated in Figure 6. This framework is utilized for both use case 1 and 2 of AC^3, where each scenario involves the streaming of data to their respective applications.

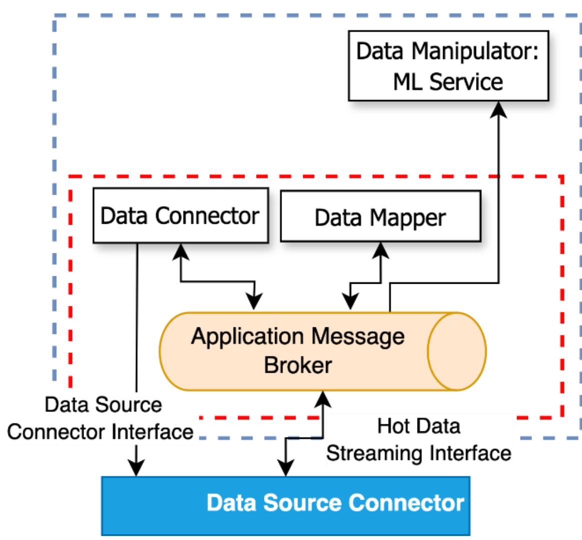

Figure 6: Structure of an AC3 Hot Data Application implementing a ML Service

[6] RabbitMQ 3.13 Documentation: https://www.rabbitmq.com/docs

[7] Kafka Documentation: https://kafka.apache.org/24/documentation.html

The process begins with the Data Connector, a key component tasked with initiating data retrieval. Based on the application's specifications and configuration parameters, the Data Connector interacts with the Data Source Connector to start the data streaming to the Application Message Broker. This broker acts as a crucial communication hub, ensuring that data generated at the source is accurately transmitted to the AC³ application via the hot data streaming interface, primarily based on the AMQP protocol. However, alternative protocols can be implemented if necessary.

Following this, the Data Mapper receives the incoming data via the Application Message Broker. Its primary function is to transform the data into a format comprehensible to the application's internal components, known as Data Manipulators. In addition to format conversion, the Data Mapper may also perform minor processing or aggregation tasks, such as time-based analytics or data aggregation.

Lastly, the Data Manipulator, which is integral to the application's business logic, integrates the data into the application's operational workflows. The data manipulator is essentially the interface with the core application that uses the data to perform certain tasks.

5.2.2 Cold Data management application addons

In a Cold Data setting the AC³ application is expected to look like the one presented in Figure 7. This is the simplest case, with only two addons needed: the **Data Connector** and the **Shared Storage**.

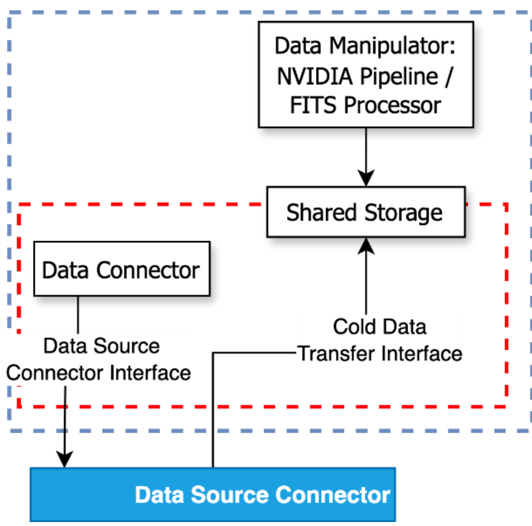

Figure 7: Structure of an AC3 Cold Data Application processing stored Video or FITS files

The Cold Data Retrieval begins with the Data Connector. Acting on instructions defined within the application's deployment parameters, the Data Connector interfaces with the Data Source Connector to initiate and orchestrate the transfer of the cold data to the Shared Storage of the application. Its interface, serving as a communication bridge, ensures that the data retrieval requests are accurately relayed to the Data Source Connector of the external data repositories. These repositories, where cold data resides, are accessed to fetch the required datasets. Then, the Cold Data Transfer Interface oversees the secure and efficient transfer of cold data from the repositories to the application's domain, ensuring data integrity and confidentiality. Upon successful transfer, the cold data is temporarily staged in Shared Storage, a storage location accessible by the application's components where the Data Manipulator has access. The Data Manipulator, that is dependent on the application's business logic is then poised for integration of the data into the application's operational workflows.

With the cold data being processed and ready, the addons are integrated into the application, enriching the application's functional capabilities and enhancing its decision-making processes. This integration marks the culmination of the cold data handling process, where the once dormant data now actively contributes to the application's dynamic environment, driving insights, and fostering informed decisions.

Throughout this process, monitoring mechanisms are in place to ensure the smooth flow of operations and to address any potential issues promptly. The orchestration of these components and stages within the AC³ architecture exemplifies a holistic approach to cold data management, ensuring that cold data is not merely stored but is actively harnessed to add value to the application ecosystem.

5.3 Planned implementations and extensions for final phase

Most of the application-specific components, which we refer to as **Data Manipulators**, are already in place as integral parts of each use case application. These components will undergo adaptations to integrate seamlessly with the AC³ managed applications, tailored according to the unique requirements of each Use Case. For instance, the UCM team is currently transitioning to containerize its application, which is designed to analyse FITS files, ensuring it is fully compatible with the AC³ framework.

The **Application Message Broker**, essential for transferring hot data in real-time for use cases 1 and 2, is already operational. The solution requires some use case specific testing to verify that its performance aligns with the demands of these use cases, ensuring reliability and efficiency in data handling.

In the upcoming period, significant effort will be directed towards the development of **Data Mappers**, which are crucial for processing the specific types of data required by each use case. SPA has already laid a foundation with existing templates for multiple applications, which will serve as the starting point for our development efforts, streamlining the creation of these vital components.

Finally, significant development efforts will be directed towards implementing **Data Connectors**, based on the groundwork already established with the Data Source Connectors. This component is particularly critical as it facilitates the interaction between data sources and our system. Ensuring a flawless integration between these connectors will be a priority, as it is fundamental to maintaining the integrity and efficiency of data exchange across the entire AC³ ecosystem. This comprehensive approach to developing and refining these components will enhance the overall functionality and performance of the AC³ applications, supporting a robust and dynamic data management system.

6 Conclusions

This deliverable, D3.3, reports the core outcomes of the successfully completed initial phase of the data management implementation for the AC^3 project.

The AC^3 data management framework, encompassing the Data Catalogue, data sources, and Application Add-ons, has been effectively mapped to the overall architecture to ensure efficient, secure, and dynamic data handling. Key achievements include seamless data integration, dynamic discovery and cataloguing, advanced processing and analysis, robust security measures, and a scalable infrastructure. Future enhancements will focus on refining interfacing mechanisms, expanding functionalities, and improving security protocols. The goal for the final achievements is to position the AC^3 data management framework as a comprehensive solution for managing data across diverse and dynamic environments within the CECC architecture.

With respect to the implementation and progress of the AC^3 Data Catalogue, particular emphasis is given to the use of the Piveau platform. Key achievements include enhanced metadata management and open data compatibility achievable through the Piveau Catalogue, which fosters better data cataloguing and utilization. The integration of the Cold Data Connector, leveraging IONOS S3 services, can improve the security, scalability, and reliability of data storage. Planned future work involves refining data source interfaces, implementing Data Source Connectors based on the Eclipse Data Space Connector framework, and further developing capabilities for handling real-time data streaming and secure data sharing. These advancements position the AC^3 Data Catalogue as a robust and efficient component for managing diverse data types within the Cloud-Edge-Client Continuum, driving innovation and data-driven decision-making across the AC^3 consortium.

For the work related with data sources' interfacing solutions and data collection mechanisms, notable accomplishments include the development of Data Connectors and the implementation of the Streaming IoT Data Connector, which facilitates real-time data integration from various IoT infrastructures. Additionally, the upcoming Streaming Video Data Connector highlight advancements in real-time data processing and integration, ensuring scalability, ease of integration, and data security. Future work will focus on evaluating and expanding the capabilities of these Data Connectors, developing EDC extensions for streaming data, and refining data processing through enhanced service deployment and orchestration. These efforts will enhance the AC^3 platform's ability to handle dynamic and diverse data sources, supporting robust and efficient data management across the consortium.

Finally, for the data management application and network service components, specific addons and data source connection guidelines have been specified for supporting the use case developments. AC^3 applications use a micro-service architecture, and these addons operate accordingly, offering multiple microservices for application developers' needs. The key achievements include the deployment of components like Data Connector, Message Broker, Data Mapper, shared storage location, and one or more Data Manipulators. The adopted technologies are mainly based on microservices architecture, AMQP protocol for communication, and shared storage solutions for data management. Micro-service addons are integrated into the AC^3 application to facilitate processing of data within the application's lifecycle. The planned implementations include utilizing real-time and historical data insights for dynamic service deployment and optimization, enhancing service management frameworks with data-driven decision-making processes, and enabling continuous service improvement. Future developments aim to expand data catalogue interfacing and adopt additional data types to facilitate end users' data sharing and usage.