D.6.3 Data Management Plan - Intermediate

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	D.6.3 Data Management Plan - Intermediate		
Work Package	6		
Contractual due date	M18: 30/06/2024	**Actual submission date**	30/06/2024
Type		**Dissemination Level**	
Lead Beneficiary	SPA		
Responsible Author	Dimitrios Amaxilatis (SPA)		
Contributors	Dimitrios Amaxilatis (SPA), Nikolaos Tsironis (SPA), Themis Saradakos (SPA), Cristina Catalan Torrecilla (UCM), Ibrahim Afolabi (FIN), Hicham Aitsaid (FIN), Mohamed Mekki (EUR), Vrettos Moulos (UPR)		
Peer reviewer(s)	Apostolos Siokis (IQU)		

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	15/02/2024	5%	Initial Deliverable Structure	D. Amaxilatis (SPA) N. Tsironis (SPA)
v0.2	04/03/2024	15%	Editing Section 3	N. Tsironis (SPA)
v0.3	10/03/2024	20%	Editing Section 4	T. Saradakos (SPA)
v0.4	27/03/2024	25%	Editing Section 5	N. Tsironis (SPA)
v0.5	02/04/2024	35%	Addition of Collected – Generated Data of the AC³ Use Cases	N. Tsironis (SPA) H. Aitsaid (FIN) C. C. Torrecilla (UCM)
v0.6	26/04/2024	50%	Addition of Collected – Generated Data of the AC³ CECC	I. Afolabi (FIN) M. Mekki (EUR) V. Moulos (UPR)
v0.7	02/05/2024	70%	Editing Section 6	D. Amaxilatis (SPA)
v0.8	06/06/2024	80%	Internal Review	A. Siokis (IQU)
v0.9	06/06/2024	90%	Executive Summary	D. Amaxilatis (SPA)
v0.9	20/06/2024	100%	Final Review and Submission	C. Verikoukis (ISI)

Disclaimer

Copyright message

List of Tables

Glossary of terms and abbreviations used

Abbreviation / Term	Description
2D	Two-dimensional
AC^3	Agile and Cognitive Cloud-edge Continuum management
AI	Artificial Intelligence
API	Application Programming Interface
ARK	Archival Resource Key
CA	Consortium Agreement
CC 0	Creative Commons Zero licence
CC BY	Creative Commons Attribution licence
CECCM	Cloud Edge Computing and Control Manager
CERIF	Common European Research Information Format
DMLC	Data Management Life Cycle
DMP	Data Management Plan
DDI	Data Documentation Initiative
DOI	Digital Object Identifier
E2E	End to End
EC	European Commission
EOSC	European Open Science Cloud
ESO	European Southern Observatory
FAIR	Findable, Accessible, Interoperable, Re-usable
FITS	Flexible Image Transport System
FL	Federated Learning
GA	Grant Agreement

GDPR	General Data Protection Regulation
I2C	Inter-Integrated Circuit
IFU	Integral-Field Unit
IoT	Internet of Things
IPR	Intellectual Property Rights
LCB	Large Compact Bundle
ML	Machine Learning
MOS	Multi-Object Spectrograph
MUSE	Multi-Unit Spectroscopic Explorer
NIS Directive	Directive on security of Network and Information Systems
OpenAIRE	Open Access Infrastructure for Research in Europe
OSR	Ontology and Semantic-aware Reasoner
PaaS	Platform-as-a-Service
PID	Persistent Identifier
PM	Particulate Matter
pPXF	Penalized Pixel-Fitting
SPI	Serial Peripheral Interface
SWRL	Semantic Web Rule Language
UAV	Unmanned Aerial Vehicle
UC	Use Case
USB	Universal Serial Bus
VLT	Very Large Telescope
VPH	Volume Phase Holographic
WCS	World Coordinate System

WP	Work Package
XAI	Explainable Artificial Intelligence

1 Executive Summary

The Intermediate Data Management Plan (DMP) for AC³ is a crucial milestone designed to ensure effective, secure, and ethical data management practices throughout the project's lifecycle. Building on the Initial Data Management Plan, this document emphasizes adherence to the FAIR principles—Findability, Accessibility, Interoperability, and Reusability—while addressing data security and ethical considerations and outlines the AC³ project's commitment to the FAIR data principles, ensuring data is findable, accessible, interoperable, and reusable. It establishes appropriate metadata standards, identifies suitable repositories, and sets conventions and templates for data generation and management. In this document we also provide a comprehensive overview of the data management lifecycle, detailing stages from data collection and generation to processing, storage, analysis, sharing, and eventual archiving or deletion. This structured approach helps us ensure proper data handling at each stage to minimize risks and maximize value.

We also provide a detailed look at the types and formats of data collected and generated across different use cases and components of the AC³ project. This encompasses information on the data collected and generated at all levels of the AC³ project from cloud, edge, and far edge platforms, as well as data related to the project's Use Cases and individual components. In addition to the information presented in the Initial Data Management Plan of the project, based on the evolution of the AC³ CECCM architecture we also provide more information on the types and nature of the data collected and produced by the different components of the CECCM that are crucial for its operation and success. This data is either datum that describe the infrastructure monitored and controlled by the CECCM, the applications deployed and managed by the CECCM, the datasets offered by the CECCM and the users using the CECCM of AC³.

Data security and privacy are emphasized as in the previous version of this document, with measures such as encryption, access controls, regular audits, and secure storage practices outlined. Privacy considerations include data anonymization techniques to protect personal information. Finally, ethical aspects of data management are also addressed, ensuring compliance with relevant laws, regulations, and guidelines. Procedures for obtaining informed consent, managing data from vulnerable populations, and addressing potential risks or biases in data collection and analysis are detailed.

Designed as a dynamic document, the DMP will evolve throughout the project's duration. Regular updates and adaptations will ensure ongoing effectiveness in promoting transparent and collaborative data management practices. By implementing this comprehensive DMP, the AC³ project demonstrates its commitment to responsible and transparent data management, enabling efficient data utilization while safeguarding privacy, security, and ethical standards.

2 Introduction

The Data Management Plan (DMP) is essential for ensuring the efficient handling, security, and ethical use of data within a project or organization. This plan covers several key areas, including open data principles, the data management lifecycle, types of data collected for specific use cases, data security and privacy measures, and ethical considerations related to data protection.

Open data is increasingly important in today's digital world, advocating that certain data types should be publicly accessible to promote transparency, collaboration, and innovation. This DMP will detail AC³'s commitment to open data principles, specifying which data will be openly accessible and the conditions for access. The data management lifecycle provides a structured approach to managing data from creation to preservation or disposal. This document describes the stages of this lifecycle, including data collection, processing, storage, analysis, sharing, and eventual archiving or deletion, ensuring proper data handling at each stage to minimize risks and maximize value.

Different use cases within AC³ may require various types of data collection. This document offers a detailed overview of the data types collected by each use case, including their sources, formats, and potential interdependencies, to illustrate the scope and diversity of the project's data.

Data security and privacy are paramount to protecting sensitive information and maintaining stakeholder trust. This DMP outlines the measures and protocols in place to ensure data security, such as encryption, access controls, regular audits, and secure storage practices. It also addresses privacy considerations, including anonymization, to protect personal information. Ethical considerations are a crucial part of the DMP, particularly regarding data protection, ensuring compliance with relevant laws, regulations, and ethical guidelines. This document outlines procedures for obtaining informed consent, managing data from vulnerable populations, and addressing potential risks or biases in data collection and analysis.

By implementing a comprehensive DMP, the AC³ project demonstrates its commitment to responsible and transparent data management practices, enabling the efficient use of data while safeguarding privacy, security, and ethical standards.

2.1 Mapping AC³ Outputs

The purpose of this section is to map AC³ Grant Agreement (GA) 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			
D6.3 Data Management Plan - Intermediate Intermediate Data management plan			
TASKS			

Task T6.3 Exploitation, Data and IPR Management	AC³'s Data Management Plan (DMP) will be developed with information related to the types of data the project will generate and collect, the standards that will be used to represent the data during the project and how partners might exploit the data resulting from the project. In addition, the DMP will include a data protection impact assessment of AC³'s requirements that will be used to ensure the project takes a data protection-by-design approach in accordance with the GDPR, EC's guidelines as well as any national and international legislation and ethics procedures applicable. The first version of the DMP will be made available in M6 to pave the way for further developments and updated twice on M18 and M36 as the project continues. Output: DMP (initial, intermediate, final) Contributors: [All partners except CDS, RHT, UCM]	3 - Open Science and Open Access 4 - Data Management Plan Overview 5 - Data Security and Privacy 6 - Ethical Considerations and Data Protection	The deliverable describes the Data Management Plan of the AC³ project. Therefore, in this deliverable the reader will find: • Open Data and Open Access principles followed • Data Management Lifecycle • Data Types collected by project and Use Cases • Data Security and Privacy principles • Ethical Considerations on Data protection

2.2 Updates with regards to D6.2

The intermediate version of the Data Management Plan (D6.3), expands upon the initial plan D6.2 by updating and detailing the data management strategies to accommodate the development and progress of the project. This adaptation is necessary to reflect the operational advancements and insights gained as the project progresses. Specifically, D6.3 revises the descriptions of the data types and collection methods to reflect the practical experience and insights gained during the project's execution phase. This includes adjustments in data handling procedures to better align with the operational realities and technological advancements encountered by the project teams.

Additionally, D6.3 provides a more nuanced approach to the lifecycle management of data, updating how data is to be collected, stored, and utilized across different stages of the project. These updates focus on refining data categorization and storage practices to ensure data integrity and accessibility, addressing any gaps identified in the initial planning stages. Such enhancements are designed to optimize the management of project data in response to evolving project needs and objectives, ensuring a more effective and efficient data management process.

This strategic update in D6.3 ensures that the data management approach not only aligns with the current state of the project but is also robust enough to accommodate future developments. By continuously refining these practices, the AC³ project ensures that data management remains a cornerstone of project success, supporting all related activities with an efficient, secure, and highly functional data infrastructure.

2.3 Deliverable Overview and Report Structure

This deliverable is divided into seven sections:

- Section 1: Executive Summary
- Section 2: Introduction – provides introductory information about the DMP, the context in which it has been elaborated as well as about its objectives and structure
- Section 3: Open Science and Open Access – describes the FAIR Data Management procedures, presents all information regarding the AC^3 promise to support research open access, the Open Research Data Pilot, and the publishing infrastructure for Open Access as well as the publishing process
- Section 4: Data Management Plan Overview – provides an overview of the Data Management Plan, a summary of the data that will be collected / generated during the activities of AC^3 including the purpose of their collection / generation as well as their types and formats and finally details the resources used for data management in AC^3 and identifies the data management responsibilities
- Section 5: Data Security and Privacy – outlines the data security strategy applied within the context of AC^3 along with the respective secure storage solutions employed
- Section 6: Ethical Considerations and Data Protection – addresses ethical aspects as well as the other relevant considerations pertaining to the data collected / generated during the project implementation
- Section 7: Conclusions

3 Open Science and Open Access

3.1 FAIR Data Principles

The Guidelines on Data Management in Horizon Europe highlight the paramount importance of ensuring the Findability, Accessibility, Interoperability, and Reusability (FAIR) of data generated through funded projects. These guidelines aim to establish robust data management practices that enable effective handling of research data. Specifically, adherence to FAIR principles involves employing standardized formats and metadata to enhance data discoverability, clearly defining data sharing protocols and determining which data will be openly accessible. Also, it encourages open repositories for data exchange and focuses on facilitating data reusability. Considering these objectives, the subsequent sections of the Data Management Plan (DMP) outline the approach employed within the AC^3 framework, encompassing strategies for achieving data findability, accessibility, and interoperability, as well as ensuring data preservation and open access to maximize its potential for reuse.

3.1.1 Findable

Any open datasets produced by AC^3 will be accompanied by data that will facilitate their understanding and reuse by interested stakeholders. These data may include basic details that will assist interested stakeholders to locate the dataset, including its format and file type as well as meaningful information about who created or contributed to the dataset, its name and reference, date of creation and under what conditions it may be accessed. Complementary documentation may also encompass details on the methodology used to collect, process and/or generate the dataset, definitions of variables, vocabularies, and units of measurement as well as any assumptions made.

The findability of the data will be achieved as follows:

- The data will have Persistent IDentifiers (PIDs) (e.g., Digital Object Identifiers (DOIs)), which are important because they unambiguously identify the data and facilitate data citation as they will be deposited in trusted repositories (e.g., Zenodo EU Open Data Portal and EOSC).
- Data will have rich metadata that will support findability, citation, and reuse. Rich metadata will provide important context for the interpretation of the data and make it easier for machines to conduct automated analysis. Standard metadata schemes (e.g., Dublin Core, CERIF, and DDI) will be followed.
- Data made available will follow specific naming conventions in order to help researchers track their origin and the showcase where in the project they were generated, e.g., AC3-[WP]-[title]-[VERSION]-[DATE].[TYPE].
- Each dataset released will be accompanied with a version number so that users can identify newer versions of it and the changes between them through defined change lists.

3.1.2 Accessible

The shared data will be deposited through an open data repository, made available through the AC^3 website [[1] an external dedicated service (such as re3data, Zenodo, DRYAD, and Harvard Dataverse). However, sensitive data that will be managed during the project may need to be totally or partially opted-out of some as they will be incompatible with the need for confidentiality in connection with security issues and with existing rules concerning the protection of personal data [[2]], [[3]]. The general principles for handling Knowledge and IPR within AC^3 will be settled in the GA and Consortium Agreement (CA). These principles are in line with Horizon Europe IPR recommendations. Background and foreground results will be clearly identified in detail within the CA and when applicable, granting access rights will be clearly specified. These result lists will also be re-evaluated by the consortium regularly and updated in as a running list.

3.1.3 Interoperable

The produced data will use common, standardized and non-proprietary formats and standards and community agreed schemas, controlled vocabularies, keywords, thesauri, or ontologies where possible to be interoperable and be integrated with other data, applications, and workflows. Additionally, the project will investigate the option to deliver in a timely manner standards, specifications and methodologies stemming from project activities to ensure the maximum interoperability between the services and tools produced.

3.1.4 Reusable

The generated data will be well-documented and will have clear licensing and provenance information. README files will be used for ensuring that the data can be correctly interpreted and re-analysed by others. Such files will include amongst other information:

- Short descriptions of the included data
- For tabular data: definitions of column headings and row labels, data codes (including missing data) and measurement units
- Any data processing steps that may affect interpretation of results
- A description of what associated datasets are stored elsewhere, if applicable
- Contact information. Referring to the license issues, data will have a clear license to govern the terms of its reuse (e.g., Public Domain, Attribution, Non- commercial, No Derivatives, or other).
- List of applied data quality assurance processes followed for the collection and the sharing of the data in question.

3.2 AC³ OpenAIRE Compatibility

AC³ data will be made open and offered in compliance with the FAIR data principles and the Open Research Data Pilot (OpenAIRE) by depositing project-related research outputs, such as publications and datasets, in relevant repositories (Arxiv, Zenodo, Kaggle, etc.). This will enable us to comply with funder requirements for open access to research outputs, increase the visibility of our research, and enable reuse by other researchers.

To facilitate the deposit of research outputs, AC³ will ensure that all project-related publications and datasets are assigned persistent identifiers, such as DOIs or Archival Resource Key (ARKs). We will also ensure that all research outputs are properly licensed to enable reuse by others. Additionally, results of the project will be shared with other researchers in platforms like Zenodo, to facilitate their availability.

3.3 Publishing Infrastructure for Open Access

Consortium partners are committed to use the Open Research Europe open access publishing platform [[4]] for scientific articles to enable rapid publication times and publication outputs that support research integrity, reproducibility, transparency and enable open science practices. To ensure open access to the deposited publications, consortium partners will be free to choose between self-archiving (“green” Open Access) and open access publishing (“gold” Open access). In the first case, consortium partners will deposit the final peer- reviewed manuscript in a repository of their choice, ensuring open access to the publication within a maximum time-period of six months. Alternatively, publications in open access journals will be pursued or in journals that also offer the possibility of making individual articles openly accessible. This strategy is directly related to the “Open” paradigm that will be used for publishing project results.

3.4 Publishing AC³ Open-Source Software

The AC³ project has established an organization on Github[1] to provide access to tools and software developed and released by project partners. In most cases these tools will be released as open-source solutions that other developers and projects can use to benefit from our own work, fostering a community-driven approach to all aspects of the CECCM AC³ initiative. In this repository, partner developers and users of the AC³ platform will be able to find documentation and ask questions regarding potential issues that will be identified. Additionally, contributions to other open-source projects like EDC will be made available through this organization to help collect all our contributions in a single point to further extend the project's reach. In this manner, multiple development teams have been established, as well as multiple repositories based on the software components under development.

3.5 Publishing Process

The publishing process within the AC³ project follows the principles of open science and open access to ensure the timely dissemination of research results. The beneficiaries of the project are committed to sharing their findings in a publicly available format, while considering restrictions related to intellectual property protection, security rules, and legitimate interests.

To facilitate open access to scientific publications, the beneficiaries will adhere to the following guidelines:

1. Deposit Publications: At the time of publication, a machine-readable electronic copy of the final peer-reviewed manuscript or the published version will be deposited in a trusted repository for scientific publications.
2. Open Access Provision: Immediate open access will be provided to the deposited publications via the repository, under the Creative Commons Attribution International Public License (CC BY) or an equivalent license. For certain formats like monographs, licenses may exclude commercial uses and derivative works.
3. Research Output Information: Comprehensive information about the research output or any other tools and instruments necessary to validate the conclusions of the scientific publication will be provided through the repository.

The beneficiaries will retain sufficient intellectual property rights (IPR) to comply with open access requirements. Metadata of the deposited publications will be openly accessible, adhering to the FAIR principles and the Creative Common Public Domain Dedication (CC 0) or equivalent, providing essential details such as author(s), publication venue, Horizon Europe or Euratom funding, and persistent identifiers.

Regarding research data management, the beneficiaries will ensure responsible handling of digital research data generated within the project. Actions include apart from this running DMP document:

1. Data Depository: The data will be deposited in trusted repositories within the designated timelines, which may be federated in the European Open Science Cloud (EOSC) as per EOSC requirements.
2. Open Access to Data: Open access to the deposited data will be provided through the repository, under the CC BY or CC 0 license, following the principle of "as open as possible as closed as necessary." Justifications for not providing open access to certain data will be documented in the DMP.

[1] https://github.com/EU-AC3

3. Information for Reuse and Validation: Information about research outputs and tools needed to re-use or validate the data will be provided via the repository.

The beneficiaries will also comply with any additional open science practices and obligations imposed by the call conditions. Furthermore, a plan for the exploitation and dissemination of results, including communication activities, will be developed and regularly updated unless excluded by the call conditions.

By adhering to these publishing processes, the AC^3 project aims to promote open access to scientific publications and foster responsible research data management in line with the principles of open science.

4 Data Management Plan Overview

The DMP serves as a comprehensive roadmap for ensuring the efficient and responsible management of research data throughout the lifecycle of our project. This overview provides a high-level summary of our data management practices, highlighting our commitment to promoting data integrity, accessibility, and the long-term usability of valuable research outputs.

4.1 Data Management Life Cycle

The Data Management Life Cycle (DMLC) within the AC^3 project plays a crucial role in facilitating the efficient handling of data throughout the CECC infrastructure. The project aims to unify and federate cloud and edge resources, catering to emerging applications that require low latency, deal with large volumes of data, and utilize diverse data sources. The CECCM (Cloud Edge Computing and Control Manager) serves as the key component responsible for managing the life cycle of applications and the federated infrastructure resources, encompassing IT, networking, and data.

To achieve seamless data management, the AC^3 project leverages semantic and ontology techniques to provide context-awareness to the CECCM. By intertwining data sources, applications, user requests, and the CECC infrastructure, the CECCM ensures an optimized and harmonized application behaviour while reducing end-to-end execution time and maximizing local bandwidth utilization. Additionally, the AC^3 project addresses security and trust as inherent components of the federated infrastructure.

The data management module, integrated as a Platform-as-a-Service (PaaS) within the CECCM, is designed to streamline the application workflow. It incorporates various components such as data indexing, searching and retrieval, parsing, storing, transferring, managing, monitoring, and streaming. The module simplifies the collection and management of data sources, enabling application developers to interact with the module through Application Programming Interfaces (APIs). It facilitates data discovery, ensuring comprehensive tracking of available data sources, their formats, sensitivity, expiration dates, and geographical locations. The data management module also considers the storage policy defined by the application developer, offering flexibility in data storage options.

The AC^3 project also focuses on AI-based mechanisms for efficient application life cycle management and resource management within the CECC infrastructure. Through Machine Learning (ML) algorithms, the project builds application profiles that enhance contextual understanding and aid in decision-making processes such as initial micro-service placement, runtime management, resource scaling, and network resource updates. The AI-based resource management module predicts infrastructure resource utilization and incorporates Federated Learning (FL) models to ensure data confidentiality and security. Furthermore, the project emphasizes the interpretability of ML models to foster trust and improve decision-making, utilizing modules such as explainable AI (XAI) and machine reasoning.

In summary, the DMLC within the AC^3 project encompasses a comprehensive approach to facilitate efficient handling, integration, and utilization of data sources. By leveraging semantic and ontology techniques, AI/ML mechanisms, and a federated infrastructure, the project aims to optimize application behaviour, enhance resource management, and ensure secure and trustworthy operations within the CECC infrastructure.

4.2 Data Collection and Generation

In the AC^3 project, data collection and generation play a crucial role in achieving its objectives. The project employs various methods and technologies to collect and generate data relevant to the CECC environment. This section outlines the purpose of data collection and generation and provides an overview of the AC^3 platform types and formats of the collected/generated data.

4.2.1 Purpose of Data Collection and Generation

The primary purpose of data collection and generation in the AC³ project is to enable advanced analytics, machine learning, and artificial intelligence techniques for effective management and optimization of the Cloud Edge Continuum infrastructure. By gathering data from various sources within the infrastructure, AC³ aims to gain insights into resource utilization, performance metrics, energy consumption patterns, and application behaviours. This data-driven approach empowers the system to make informed decisions, dynamically adapt to changing conditions, and deliver enhanced services to end-users.

Furthermore, data collection and generation support research and development activities in the AC³ project. The collected/generated data serve as a valuable resource for studying and analysing the behaviour of distributed applications, evaluating the performance of different infrastructure components, and developing new algorithms, models, and optimization techniques.

The AC³ project collects and generates data from diverse platforms within the Cloud Edge Continuum infrastructure as presented in Table 2.

Table 2: Descriptions of the data collected across the project's Cloud Edge Continuum infrastructure.

Platform	Description
Cloud	Data collected/generated from the centralized cloud resources, which encompass data centers and computing clusters. This data may include information about resource utilization, network traffic, application performance, and user interactions.
Edge	Data collected/generated from the edge devices and infrastructure deployed at the network edge. This includes data from edge servers, gateways, routers, sensors, and IoT devices. The collected/generated data from the edge platform provide insights into edge resource availability, latency, bandwidth, and environmental conditions.
Far Edge	Data collected/generated from the far edge devices located in remote areas (sensing devices, drones, and other edge resources). The collected/generated data from the far edge platform enable analysis of resource availability, connectivity, and performance in remote locations.

The collected/generated data in the AC³ project can take various formats, including structured and unstructured. Structured data are organized in a well-defined format such as databases, spreadsheets, or tables, enabling efficient storage, retrieval, and analysis. Unstructured data do not have a predefined structure, such as text documents, multimedia files, or raw sensor readings. Unstructured data requires advanced processing techniques for extraction and analysis.

The AC³ project ensures that the collected/generated data are handled in compliance with applicable data protection and privacy regulations. Stringent security and privacy protocols are in place to safeguard the confidentiality, integrity, and availability of the collected/generated data throughout their lifecycle.

4.2.2 Collected – Generated Data of the AC³ Use Cases

4.2.2.1 Use-Case 1 (UC1) Types and Formats of Collected / Generated Data

UC1 primarily deals with IoT **time-series data** generated by on-site sensors monitoring the environmental and air quality conditions as well as the operational properties of equipment installed and used in the monitored

facility. IoT enabled sensing infrastructures to have been developed and can be deployed at facilities or locations around the world, such as factories, self-driving cars and low-latency communication scenarios. The UC1 testbed, provided by IQU, is an End-to-End (E2E) Beyond-5G Experimental Platform for ultra reliable low latency communications applications, such as Factory Automation, Autonomous Driving and teleoperation. This platform requires the deployment of a 5G Mobile Core, and additional software components to enable connectivity across 5G end user devices and leverage its computational capabilities. On the cloud side, high-performance and high-computing capabilities servers handle the compute-heavy operations of the platform. These services are augmented with the cloud data analytics capabilities provided by the SparkWorks (SPA) IoT Data Analytics Engine that can process and analyse unbounded streams of data from IoT devices in the cloud. Data visualization software, like Grafana or custom user interfaces are used to showcase the information generated by both the sensing as well as the networking infrastructure. 5G-enabled end user devices are currently deployed in the testbed either in the form of smartphones (e.g., OnePlus 8T 5G phones) or in the form of single-board computers (e.g., Raspberry Pis with a SIM8200EA-M2 5G modem) or 5G gateway devices (e.g., Teltonika TRB500 5G gateway). A similar infrastructure is used by SPA to facilitate the development of the UC and to test the software of AC^3.

On the sensing side, high-quality and high-accuracy sensing devices are used to collect real-time information for the monitored environments:

- Sensirion SCD4X CO2 sensors are available in this testbed collecting information about the **temperature**, **relative humidity** and **CO_2 concentration**;
- Bosh BME68X sensor are available for monitoring environmental information including **temperature**, **relative humidity** and **air quality index**;
- PlanTower PMSX003 sensors can be used to monitor the **particulate matter** (PM) levels of the facility;
- SPH0645 mems microphone sensor can also be installed to generate data regarding the **noise levels** of each monitored location.

All sensing devices are compatible with edge devices and can communicate with them through interfaces including USB, I2C or SPI for fast sampling and accurate measurements.

Additional data can be generated based on the sensed parameters mentioned above, such as the **occupancy** of specific areas (e.g., based on CO_2 concentration, noise, or power usage). To do so the aforementioned data can be processed at the edge and be combined with additional **metadata** regarding the structure of the facility monitored including **building information**, **room size**, **volume** and **usage type**, **expected occupants**, **time** of **day** or **day** of **year**.

4.2.2.2 Use-Case 2 (UC2) Types and Formats of Collected / Generated Data

UC2 leverages and combines IoT (Internet of Things), camera, and unmanned aerial vehicle (UAV) technologies to provide an efficient and powerful video surveillance system. The IoT devices can be deployed on the ground or onboard UAVs to gather valuable monitoring data, such as CO, CO_2, and passive infrared sensors (PIR). For instance, a PIR could be deployed on top of a UAV that triggers the camera onboard only if a moving object is detected. Other sensors, such as temperature or door/window sensors, can provide extra contextual information that enhances the quality of the video surveillance system. The use of UAVs will empower the video surveillance system with the capacity to cover large areas and the ability to omit surveillance blind spots. In fact, the UAVs will help cover blind spots not covered by the fixed cameras due to their locations.

In this use case, we target a heterogeneous system consisting of different IoT devices and cameras with different computational capabilities. Some IoT devices and cameras have more computation capacity for performing built-in processing capabilities for cutting-edge functionalities. The built-in processing helps diminish network overhead and increases data privacy. By leveraging machine learning, more precisely deep learning, the video

surveillance system can detect and track objects, recognize faces, and detect anomalies or suspicious behaviours. According to the computational capacity of IoT devices and cameras on the ground or onboard UAVs, the data analytics of video contents and sensor data can happen **locally** or **offload** to dedicated **servers or devices**. If the live video contents are pre-processed locally in the camera, a **post-processed video stream** and its **metadata** will be sent to the dedicated servers. Otherwise, a **raw video stream** will be offloaded to another IoT device, camera, or server for preprocessing before streaming the post-processed video stream and its metadata to the dedicated servers. Regarding sensor data processing, instead of sending **raw data** to dedicated servers, pre-processing (e.g., Kalman and Particle filters) can happen locally or offload to more powerful devices for detecting correlation and making more accurate decisions.

We plan to leverage micro-services with their replications offered by AC^3 to ensure the high availability and reliability of the video surveillance system. We plan to split and offload the same task into multiple devices to ensure the real-time processing of **IoT sensor data** and **video streams**. For instance, data analytics of video content via deep learning will be distributed by leveraging federated learning and transfer learning mechanisms. We expect to distribute the deep learning models (e.g., YOLOv7) into multiple containers that can be used either for inference or fine-tuning previously trained models.

Based on the above, we expect the following type of data:

1. Time-series data that serves to monitor and manage different UAVs.
2. Time-series data related to resource utilization and operation statistics of IoT devices, UAVs, and cameras.
3. Time-series sensor data that can be either raw or pre-processed sent by different IoT devices.
4. Raw video stream streamed from the devices and cameras to the dedicated servers.
5. Post-processed video stream and its metadata (object type, location, id, size, direction, speed, colour, time) are treated at the devices and cameras before being streamed to the dedicated servers.
6. Machine learning parameters that are periodically sent in federated learning for fine-tuning deep learning models and inferences.

4.2.2.3 Use-Case 3 (UC3) Types and Formats of Collected / Generated Data

The primary data format utilized in UC3 is **datacubes**. A datacube comprises a three-dimensional array of data values, incorporating two spatial axes that represent a specific region of the celestial sphere containing the targeted galaxy, as well as a spectral axis indicating wavelength or frequency. These datacubes are stored digitally in a **Flexible Image Transport System (FITS)** format, specifically designed for astronomical data. Some common metadata found in the FITS data cube include:

- **Header**: it contains key information about the data cube, such as the observing date and time, instrument configuration, exposure time, and observational parameters;
- **WCS (World Coordinate System) information**: it provides the necessary information for spatial and spectral mapping of the data cube. It includes details about coordinate systems, pixel scales, reference points, and coordinate transformations;
- **Observation parameters**: metadata related to the observational setup, such as the observing mode, integration time, telescope pointing, and instrument settings;
- **Object identification**: it may include details about the observed object, such as its name, coordinates and other relevant identifiers. This information assists in associating the data cube with a specific astronomical source;
- **Data quality and flags**: metadata regarding data quality indicators, error estimates, and data flags may be present. These flags can indicate issues such as cosmic-ray hits, bad pixels, or regions affected by instrumental artifacts;

- **Observing log**: additional metadata may include observing log information, which records details of any manual interventions or other relevant notes taken during the observing session.

As part of the collected data, we are working with MEGARA and MUSE datacubes.

MEGARA is an optical Integral-Field Unit (IFU) and Multi-Object Spectrograph (MOS) that has been specifically designed for the 10.4m GTC telescope located in La Palma, Spain. Within the scope of this project, we are utilizing the Large Compact Bundle (LCB) mode, an IFU configuration that covers a 12.5 arcsec x 11.3 arcsec area on the sky, with each spaxel measuring 0.62 arcsec. Our data collection process involves acquiring observations with intermediate-to-high spectral resolutions (R~6,000, 12,000, and 18,700), corresponding to LR, MR, and HR spectral setups, respectively. To cover a wide wavelength range spanning from 3700 to 9800Å, we employ different Volume Phase Holographic (VPH) gratings. These gratings enable efficient dispersion and facilitate the acquisition of spectroscopic data for our project.

MUSE (Multi Unit Spectroscopic Explorer) is an integral field spectrograph developed for the Very Large Telescope (VLT) of the European Southern Observatory (ESO). MUSE covers a relatively large field of view spanning approximately 1 square arcminute, allowing the study of extended structures such as galaxies in their entirety. MUSE employs an IFU to simultaneously capture spatial and spectral information across its field of view, providing detailed spectroscopic data for every pixel. With a spectral resolution of up to 3000, it can discern fine spectral features crucial for understanding the dynamics and chemical composition of celestial objects.

In addition to scientific observations, spectrophotometric observations of standard stars are necessary to establish a response function for calibrating the absolute flux scale. At the beginning and end of each night, arc calibration lamp frames and twilight sky flat-field/lamp flat-field data are also obtained, respectively. The reduction of the raw science data follows a series of standard procedures, including bias subtraction, cosmic-ray removal, flat-fielding, tracing and extracting the spectra, applying arc calibration solutions, performing sky-subtraction, and conducting flux calibration. This reduction process is carried out using the established pipeline.

As part of the generated data, we will obtain the following information:

- Two-dimensional (2D) maps using FITS format:
    - ***kinematic parameters*** (also known as Gauss-Hermite moments) such as the line-of-sight velocity, velocity dispersion and higher order velocity moments (e.g., h3 and h4). The previous data describe the motion and dynamics of the observed galaxy. These values would be provided by the pPXF spectral synthesis software.
    - ***light-weighted ages*** to recover information about the stellar populations of each galaxy. These data would be provided by different software's (pPXF, STARLIGHT and STECKMAP).
    - ***stellar metallicities*** to know the abundances of metals. These data would be provided by different software's (pPXF, STARLIGHT and STECKMAP).
    - ***extinction maps*** to recover the spatial distribution of the amount of light attenuation caused by interstellar dust. These data would be provided by different software's (pPXF, STARLIGHT and STECKMAP).
- Datacubes using FITS format for the case of the pPXF spectral synthesis software:
    - ***best-fit model*** that contains the information about the synthetic spectrum that best matches the observed spectrum in each spaxel. It represents the combination of stellar templates and kinematic parameters (such as velocity and velocity dispersion) that minimizes the differences between the observed and model spectra.
    - ***residuals*** which show the representation of the differences between the observed spectrum and the best-fit model spectrum in each spaxel of the datacube. It shows the discrepancies between the observed data and the model prediction at each wavelength or spectral pixel. The residual map helps to identify regions where the model may not accurately reproduce the observed features.

The previously generated data could be done spaxel-by-spaxel or using the Voronoi binning technique which is a method used in astronomical data analysis to group neighboring spaxels together into spatially coherent regions based on a signal-to-noise criteria.

4.2.3 Collected – Generated Data of the AC³ CECC

The AC³ CECC during its operation and activities needs to collect data regarding its operation and the operation of the deployed and managed services. These range from **service operation statistics**, **device utilization statistics**, **network traffic** or **communication statistics,** as well as **performance statistics** for the user applications deployed and managed. These statistics are expected to be mostly in the form of time-series data collected in real-time based on the operation of the system. The same data will also be used by the core AC³ services to allow for the better placement of services in the different of the AC³ platform as well as the lifecycle management over the CECC infrastructure. Additionally, metadata like the specification of the various CECC infrastructure devices or compatibility information will be collected and needed for the operation of AC³ . This data will be needed for the better assignment of the services and ensuring the decisions of the CECCM are correct and will not affect the operation of the user's applications. Each component of the CECC is responsible for collecting, handling, storing and safeguarding all the data collected in order to facilitate the operations of the whole CECC.

4.2.3.1 Collected – Generated Data of the AC³ CECC Ontology and Semantic-aware Reasoner

The data ingested and produced by the AC³ CECC Ontology and Semantic-aware Reasoner (OSR) can be elucidated as follows:

Collected Data:

- **Policies**: Formulated by users, these are comprehensive sets of principles and directives that dictate the operation of microservices and the CECC infrastructure. For example, a policy might stipulate that a microservice should scale up when CPU usage exceeds 80%.
- **SWRL Rules**: Consisting of specific conditions and actions, these rules are expressed in the Semantic Web Rule Language (SWRL) and serve to delineate the criteria for policy validation. An example of a SWRL rule could be "If CPU usage > 80%, then trigger scale-up action."
- **Ontology Information**: This includes detailed descriptions of classes, subclasses, and object properties that define the domain of microservices policies. This may involve concepts such as microservice (representing individual services), condition (defining prerequisites for policy implementation), metric (categorizing performance measures like CPU usage), operator (specifying comparison types like 'greater than'), threshold (determining the limit value for a metric), and action (defining responses like 'scale up').
- **Contextual Data**: This category encompasses metadata related to CECC infrastructure devices, operational statistics of services, utilization metrics of devices, network traffic data, and performance metrics of user applications. For instance, this might include the number of requests handled by a microservice per minute or the average response time of a service.

Generated Data:

- Structured Policies in Ontology: Policies are methodically structured and depicted within the ontology, ensuring clarity and systematic organization of all pertinent information and relationships. This structured representation facilitates the interpretation and application of policies within the CECC infrastructure.
- Validated Policies: Utilizing the reasoner, the OSR evaluates the policies against the SWRL rules and the ontology. This assessment verifies the policies' adherence to the predefined criteria, ensuring their

compliance, consistency, and validity. For example, a policy might be validated to ensure it does not conflict with existing policies or system constraints.
- Deduced Knowledge: Through the application of SWRL rules to the ontology, the reasoner derives new insights, pivotal for informed decision-making and the optimization of the CECC infrastructure's operations. This deduced knowledge might include identifying potential bottlenecks in the system or suggesting areas for resource allocation.
- Feedback on Policy Validity: The OSR provides immediate feedback on the validity of policies, enabling users to make requisite adjustments to ensure compliance. This feedback mechanism ensures that policies are continually refined and aligned with the system's operational requirements.
- Semantic Validation Outcomes: The OSR confirms that the policies are semantically coherent and aligned with the business logic outlined by the SWRL rules. This validation ensures that policies are logically sound and applicable within the context of the CECC infrastructure.

4.2.3.2 Collected – Generated Data of the AC³ CECC AI-based LCM

The AI-based LCM is responsible for mapping applications to infrastructures, it uses the AI-based application profile and infrastructure profile to decide about the initial placement of applications. It also manages the applications runtime including termination, migration resources scaling. The data collected and generated by the AI-based LCM is the following:

Collected data:

- Application descriptors: The AI-based LCM receives from the OSR the descriptors of the applications that will be deployed on the CECC infrastructure.
- Application profile data: the information about the applications requirements and predicted future resources usage.
- Infrastructure profile data: Information on the infrastructure resources, availability and carbon footprint
- AI models: The models used to decide on the initial application placement, application migrations, applications resource scaling.

Generated data:

- Stats on applications: List and number of applications that are waiting to be deployed, deployed, terminated, deleted.
- Stats on the operations: List and number of instantiations, deletion of applications and their timestamp, number of times an application has been migrated or scaled and their timestamp.
- Stats on infrastructures: Number of applications on each infrastructure, list and number of infrastructures used, discovered and deleted(unavailable).
- The AI-models input and output: The LCM can provide the input and output of the used AI models and XAI algorithms.

4.2.3.3 Collected – Generated Data of the AC³ CECC Adaptation Gateway

The adaptation Gateway is responsible for creating the adaptation agents and contributing to infrastructure monitoring based on information obtained from the discovery module.

Collected data:

- Newly integrated infrastructures: The Adaptation gateway receives requests from the decision enforcement to create adaptation agents for the newly added infrastructures.
- LMS information: The list of available LMS and their data such as the LMS API and monitoring endpoints from the Resource Discovery module.

Generated data:

- Adaptation agents' stats: The list of Adaptation agents, the time of creation and deletion of each Adaptation agent.

4.2.3.4 Collected – Generated Data of the AC³ CECC LMS

The LMS is responsible for representing the component the CECCM communicates with to enforce its decisions on a certain site. Its data is related to the managed infrastructure, and it includes:

Collected data:

- The Adaptation agent requests: The time and type of requests received from the Adaptation agent of the concerned infrastructure.

Generated data:

- Infrastructure stats: The evolution of available compute and networking resources in the infrastructure.
- Operations stats: The number of usages of the LMS API, type of operations and objects created.
- Availability data: The availability of the infrastructure, downtimes.
- Carbon footprint of the infrastructure: such as the source of energy used by the infrastructure nodes and the rate of energy usage per resources consumed.

4.2.3.5 Collected – Generated Data of the AC³ CECC Application Gateway

The AC³ CECC Application Gateway is responsible for facilitating the interaction between application developers and the Cloud Edge Computing Continuum Manager (CECCM) to develop, deploy, and manage the lifecycle of applications. The data collected and generated by the Application Gateway are essential to ensure efficiency and performance of applications within the CECC framework.

Collected Data:

- Service Operation Statistics: This includes metrics of the services, such as response times, uptime, and error rates.
- Device Utilization Statistics: Data on the usage of devices within the CECC infrastructure, including CPU, memory and storage.
- Network Traffic information: Information on network usage (bandwidth usage, latency, and packet loss).
- Performance Statistics of User Applications: Metrics related to the performance of applications deployed by the CECC, such as load times and resource consumption.

Generated Data:

- Policies in Ontology: Policies structured within the ontology to ensure clarity and systematic organization of all information and relationships.
- Validated Policies: Policies evaluated by the reasoner against predefined criteria to ensure consistency, and validity.
- Feedback on Policy Validity: Feedback on policy conformity to ensure alignment with the requirements.
- Semantic Validation Outcomes: Confirmation that policies are semantically aligned with business logic, ensuring they are applicable in the CECC infrastructure.

4.3 Resources for Data Management

4.3.1 Data Management Roles and Responsibilities

In the AC³ project, effective data management relies on clearly defined roles and responsibilities to ensure the responsible handling and utilization of data throughout the project lifecycle. The following roles and responsibilities play a crucial part in the project's data management framework:

Project Coordinator (ATH/ISI): The project coordinator takes on the responsibility of overseeing and coordinating the data management activities. Ensures that the data management plan is effectively implemented, and that all data-related tasks are carried out in compliance with relevant policies and guidelines. The coordinator collaborates closely with the Work Package (WP) leader, task leaders, and UC leaders to ensure seamless communication and coordination among team members regarding data management. They serve as a central point of contact for data-related matters, facilitate data sharing and collaboration, and ensure that data management responsibilities are assigned and fulfilled by the respective partners. The project coordinator plays a crucial role in maintaining data integrity, addressing data-related challenges, and ensuring that the project's data management goals are achieved.

Work Package Leaders: The WP leader is responsible for overseeing a specific work package, which may involve multiple tasks and activities. In terms of data management, the WP leader works closely with the project coordinator to ensure that the data management plan aligns with the objectives of the work package. They collaborate with task leaders and UC leaders to define data requirements, establish data collection and storage procedures, and monitor the progress of data-related tasks within their work package. The WP leader provides guidance and support to task leaders and UC leaders in implementing effective data management practices, ensuring data quality, and addressing any issues or challenges that arise. They also liaise with the project coordinator to facilitate data integration and sharing across work packages, ensuring that data management efforts are coherent and contribute to the overall project goals.

Task Leaders: Task leaders are responsible for overseeing specific tasks. They work closely with the WP leader and project coordinator to define the data management requirements and objectives for their respective tasks. Task leaders are responsible for collecting, organizing, and storing data generated from their tasks in accordance with the project's data management plan. They ensure that data is properly documented, labelled, and preserved for future reference and analysis. Task leaders collaborate with UC leaders and other team members to facilitate data exchange and collaboration within their tasks. They also provide regular updates on the progress of data-related activities to the WP leader and project coordinator, ensuring that data management responsibilities are fulfilled within their assigned tasks.

Use Case Leaders: UC leaders are responsible for driving the use case activities within the AC³ project. In terms of data management, UC leaders work closely with the project coordinator, WP leader, and task leaders to identify the data needs and requirements specific to their use cases. They collaborate with task leaders and other team members to ensure that data collection, processing, and analysis align with the objectives of their use cases. UC leaders play a key role in validating and interpreting the data generated within their use cases, ensuring its accuracy, reliability, and relevance. They provide insights and recommendations based on the analysed data, contributing to the overall project outcomes and objectives. UC leaders collaborate with the project coordinator and other roles to ensure effective data management practices are implemented within their use case activities.

Data processors: Data processors are the project partners that collect, digitise, anonymise, store, destroy and/or otherwise process data for the specific purpose of the activity in which it has been collected/generated within the framework of the project. They are responsible for appropriately collecting the necessary consent for

processing data as well as for ensuring that the informed consent form and information sheet used to this end were properly adjusted to the needs of the activity they were participating and any particularities applicable to their organisation. They are also responsible for managing the consents they have retrieved with a view to demonstrating their compliance with the relevant applicable EU and national regulation.

5 Data Security and Privacy

Ensuring the security and privacy of data is of paramount importance within the AC³ project. This section outlines the measures and plans in place to safeguard data throughout its lifecycle, including data security planning, data preservation, archiving, and disposal, data anonymization, as well as data access and use policies.

5.1 Data Security Plan

AC³ implements a comprehensive data security plan to protect sensitive and confidential information from unauthorized access, data breaches, and cyber threats. The plan includes the following key components:

- **Access Controls**: Strict access controls are implemented to ensure that only authorized personnel can access the data. User authentication mechanisms, role-based access controls, and encryption techniques are employed to enforce data security.
- **Encryption**: Data is encrypted both at rest and in transit to prevent unauthorized access. Robust encryption algorithms are utilized to protect the confidentiality and integrity of the data.
- **Network Security**: AC³ maintains a secure network infrastructure with firewalls, intrusion detection systems, and regular security assessments to identify and mitigate potential vulnerabilities.
- **Incident Response**: An incident response plan is in place to handle security incidents and breaches promptly. This includes procedures for reporting, investigating, and mitigating security breaches, as well as notifying relevant parties as required.

5.2 Data Preservation, Archiving and Disposal Plan

To ensure the long-term preservation and responsible management of data, AC³ follows a comprehensive plan for data preservation, archiving, and disposal. The plan includes the following elements:

- **Preservation**: Data is preserved in formats that are accessible, secure, and compatible with future technologies. This includes regular backups, redundant storage systems, and periodic verification of data integrity.
- **Archiving**: AC³ maintains an archiving strategy to retain valuable data for future reference and research purposes. This involves proper documentation, metadata management, and adherence to relevant archival standards.
- **Disposal**: When data is no longer required or legally permissible to retain, AC³ follows appropriate procedures for data disposal. This includes secure deletion methods that render the data irretrievable and the documentation of disposal activities for audit and compliance purposes.

5.3 Data Anonymization

Data anonymization is a crucial aspect for every project and service developed, ensuring the protection of individuals' privacy while allowing for valuable data analysis involving the process of removing or altering identifiable information within datasets to prevent the identification of specific individuals. To achieve this, various techniques can be employed, such as removing direct identifiers like names and addresses, aggregating data to make it less granular, and generalizing certain attributes. Additionally, other potentially identifying information, such as dates of birth or unique characteristics, may be masked or replaced with pseudonyms.

In AC³ the amount of personal data collected and used are extremely limited due to the nature and context of the project's use cases. These operations are of concern mostly to UC2, due to the use of cameras for security reasons. In this context, and where needed sufficient techniques for the removal of personally identifying characteristics will be employed to comply with data protection regulations, such as the General Data Protection

Regulation (GDPR). Additional measures will be employed in the data collected from the core platform of AC³ as needed.

5.4 Data Access and Use Policy

AC³ implements a robust data access and use policy to govern the appropriate handling, sharing, and utilization of data. The policy encompasses the following principles:

5.4.1 Authorized Access

Only individuals with legitimate research or operational purposes are granted access to the data. Access rights are assigned based on role, responsibility, and need-to-know basis.

5.4.2 Data Sharing

AC³ facilitates data sharing among authorized parties through secure and controlled mechanisms. Data sharing agreements, data transfer protocols, and secure communication channels are established to ensure data protection during sharing.

5.4.3 Responsible Use

Users of AC³ data are obligated to adhere to ethical guidelines and legal requirements pertaining to data use. This includes using the data only for the specified purposes, maintaining data confidentiality, and refraining from unauthorized data re-identification or re-disclosure.

5.4.4 Monitoring and Auditing

AC³ employs monitoring and auditing mechanisms to track data access, use, and compliance. This includes regular monitoring of data access logs, activity tracking, and periodic audits to ensure adherence to the data access and use policy.

5.4.5 Data Governance

AC³ establishes a data governance framework to oversee the management, quality, and integrity of the data. This includes defining roles and responsibilities, establishing data stewardship processes, and maintaining data documentation and metadata.

5.4.6 Compliance and Enforcement

AC³ enforces the data access and use policy through regular compliance checks, audits, and appropriate disciplinary measures for policy violations. Compliance with legal and regulatory requirements, including data protection laws, is a priority within the project.

6 Ethical Considerations and Data Protection

Ethical considerations and data protection are fundamental aspects of the AC³ project. This section addresses the ethical aspects of data management, data protection, privacy concerns, and compliance with the General Data Protection Regulation (GDPR).

6.1 Ethical Aspects of Data Management

AC³ upholds ethical principles in all stages of data management to ensure the rights and interests of individuals and organizations are respected. Key ethical considerations include:

6.1.1 Informed Consent

AC³ ensures individuals have the necessary information and understanding to provide informed consent for data management activities. Transparent and clear consent mechanisms are implemented to enable individuals to make voluntary and informed decisions regarding their data.

6.1.2 Data Minimization

AC³ follows the principle of data minimization, collecting only relevant and necessary data to mitigate privacy risks and unauthorized access. Advanced techniques such as data anonymization and pseudonymization are employed to protect individual identities whenever feasible.

6.1.3 Purpose Limitation

Data collected within AC³ is used solely for the purposes specified during the consent process. Any further use of the data is in line with these defined purposes, ensuring data is not repurposed without explicit consent or legal justification.

6.1.4 Transparency and Accountability

AC³ promotes transparency by providing individuals with clear information about how their data is managed, who has access to it, and how it is used. Mechanisms are in place to ensure accountability for data management practices, including regular audits and compliance checks.

6.2 Data Protection and Privacy

AC³ prioritizes data protection and privacy to safeguard the confidentiality, integrity, and availability of data. Key aspects include:

- **Secure Data Storage**: AC³ utilizes secure storage solutions and encryption mechanisms to protect data at rest, minimizing the risk of unauthorized access or data breaches. Access controls and user authentication mechanisms are implemented to ensure only authorized individuals can access the data.
- **Data Transfer and Sharing**: When transferring or sharing data within AC³, strict protocols and encryption techniques are employed to safeguard data during transit. Secure communication channels and data anonymization techniques are used to protect the privacy of individuals involved.

6.3 GDPR

AC³ is committed to compliance with the General Data Protection Regulation (GDPR). The GDPR sets guidelines for the processing of personal data of individuals within the European Union (EU). AC³ ensures that data management practices align with GDPR requirements, including:

6.3.1 Lawful Basis for Data Processing

AC³ identifies and applies the appropriate lawful basis for data processing, such as consent, legitimate interest, or contractual necessity.

6.3.2 Individual Rights

AC³ respects the rights of individuals under the GDPR, including the right to access, rectify, erase, and restrict the processing of their personal data.

6.3.3 Data Portability

AC³ facilitates data portability, allowing individuals to obtain and transfer their personal data in a structured, commonly used, and machine-readable format upon request.

6.3.4 Data Protection Impact Assessments (DPIAs)

AC³ conducts DPIAs to assess and mitigate privacy risks associated with data processing activities that may pose a high risk to individuals' rights and freedoms.

6.3.5 International Data Transfers

AC³ ensures that any transfer of personal data to countries outside the European Economic Area (EEA) is done in compliance with GDPR provisions, such as using appropriate safeguards like standard contractual clauses or obtaining individual consent.

6.3.6 Accountability

AC³ maintains documentation, records, and policies demonstrating compliance with the GDPR. Regular audits and internal reviews are conducted to assess and improve data protection practices.

By incorporating ethical considerations and adhering to data protection regulations like the GDPR, AC³ strives to ensure responsible and trustworthy data management practices while safeguarding individuals' privacy rights.

7 Conclusions

The AC³ project's Initial Data Management Plan sets the stage for implementing a robust data management methodology. It establishes overarching methodological principles that the project partners will follow, with a focus on making the collected, processed, and generated data as FAIR as possible, while also addressing data security and ethical considerations.

The DMP provides a comprehensive overview of the valuable datasets expected to be created within the project, emphasizing the methodology for their management throughout the project's lifespan and beyond. As the project activities progress, this information will be continuously enriched and updated to ensure effective data management.

As we used FAIR data management guidelines, our DMP emphasizes the importance of ensuring data findability, accessibility, interoperability, and reusability. We have identified proper conventions and metadata to make our data findable, identified suitable repositories for accessibility, and defined steps to enhance data interoperability and reusability. Additionally, the DMP provides conventions and templates that will guide the project in managing data generation effectively.

The AC³ project's DMP is a living document that will evolve and be extended throughout the project's lifetime by all relevant partners. Regular updates and adaptations will ensure its continued effectiveness in facilitating transparent and collaborative data management practices.

The Intermediate Data Management Plan (DMP) for the AC³ project builds on the foundations established in the Initial Data Management Plan, further solidifying our data management methodology. This plan outlines the guiding methodological principles that all project partners will adhere to, prioritizing the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) while also addressing data security and ethical considerations.

The DMP provides an in-depth overview of the valuable datasets expected to be created within the project as part of the 3 project Use Cases, detailing the management methodology for these datasets throughout the project’s lifecycle and beyond. As project activities advance, this information will be continually enriched and updated to ensure effective data management. We also provide information on the data collected and generated by all other internal and user-facing components of AC³ in order to give the readers a better understanding of the amount of different data types and data collections.

Following FAIR data management guidelines, our DMP emphasizes the importance of making data findable, accessible, interoperable, and reusable. We have established appropriate conventions and metadata standards to enhance data findability, identified suitable repositories to ensure accessibility, and outlined steps to improve data interoperability and reusability. Additionally, the DMP includes conventions and templates to guide the project in effectively managing data generation.

The AC³ project’s DMP is a dynamic document that will continue to evolve and expand throughout the project’s duration with input from all relevant partners until its final version provided by the end of the project. These updates and adaptations will ensure its ongoing effectiveness in promoting transparent and collaborative data management practices.

8 References

[1] "AC³ Website," [Online]. Available: https://ac3-project.eu.
[2] European Parliament and the Council of Europe, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, 2016.
[3] European Parliament and the Council of Europe, Directive 2022/2555 on measures for a high common level of cybersecurity across the Union, 2022.
[4] "Open Research Europe," [Online]. Available: https://open-research-europe.ec.europa.eu/.