D1.2

DATA MANAGEMENT PLAN (DMP)

Revision: v 1.0

Work package	1
Task	T1.1 – T1.3
Due date	30/06/2025
Submission date	30/06/2025
Deliverable lead	UBITECH (UBI)
Version	1.0
Authors	Dimitris Manolopoulos, Dimitrios Klonidis, Georgios P. Katsikas (UBI); Arzum Koca Celiktenli (ARTHUR)
Reviewers	Ioanna Drigkopoulou (INTRA)
Abstract	This Data Management Plan (D1.2) outlines COP-PILOT's comprehensive strategy for handling all project data—spanning collection, processing, storage, sharing, preservation, and ethical governance—across its IoT–edge–cloud use cases. It summarizes the types, formats, volumes, and provenance of both existing and newly generated datasets; defines measures to ensure data are Findable, Accessible, Interoperable, and Reusable (FAIR); and describes GDPR-compliant procedures for both personal and non-personal data. Key provisions include persistent identifiers and standardized metadata schemas, secure repository deposits with clear access conditions, role-based encryption and audit trails for sensitive telemetry, and detailed provenance capture using W3C PROV-O. The plan also addresses cross-sector ethical considerations, high-risk AI deployments, and national/regional compliance requirements, and establishes governance structures to manage restricted-access data and Data Access Committees. By embedding FAIR principles, privacy-by-design, and robust security controls, this DMP ensures that COP-PILOT's valuable data assets remain ethically managed, legally compliant, and broadly reusable throughout and beyond the project's lifetime.
Keywords	Data; Data Management Plan (DMP); FAIR Principles; GDPR Compliance; IoT–Edge–Cloud Integration; Metadata Standards; Persistent Identifiers; W3C PROV-O Provenance; Privacy-by-Design; Secure Data Sharing; Role-Based Encryption; Data Access Governance

Co-funded by the European Union

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
V0.1	22/04/2025	ToC	*Dimitris Manolopoulos (UBI)*
V0.2	23/04/2025	Initial drafting of Section 1, Input to ToC	*Dimitris Manolopoulos (UBI);* *Arzum Koca Celiktenli (ARTHUR)*
V0.3	25/04/2025	Drafting Sec. 2 and updates to Sec. 1	*Dimitris Manolopoulos (UBI)*
V0.4	05/05/2025	Review and inputs to Sec. 1 & 2	*Arzum Koca Celiktenli (ARTHUR)*
V0.5	07/05/2025	Drafting Sec. 3-5	*Dimitris Manolopoulos (UBI)*
V0.6	08/052025 09/05/2025	Updating Table 2 Drafting Sec. 4	*Arzum Koca Celiktenli (ARTHUR)*
V0.7	17/05/2025	Contributions to Table 2 and review of all sections	*All WP leaders and Cluster leaders*
V0.8	28/05/2025	Peer review	*Ioanna Drigkopoulou (INTRA)*
V0.9	30/05/2025	Review comments implementation	*Dimitris Manolopoulos (UBI)*
V1.0	30/06/2025	Submission to the EC portal	*Ioanna Drigkopoulou (INTRA)*

Grant Agreement No: 101189819 | **Topic:** HORIZON-CL4-2024-DATA-01-03
Call: HORIZON-CL4-2024-DATA-01 | **Type of action:** HORIZON-IA

DISCLAIMER

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Co-funded by the European Union (COP-PILOT, 101189819). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

COPYRIGHT NOTICE

Project Co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	DMP	
Dissemination Level		
PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)	☑
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

TABLE OF CONTENTS

TABLE OF CONTENTS 5
EXECUTIVE SUMMARY 6
LIST OF FIGURES 7
LIST OF TABLES 8
ABBREVIATIONS 9
1 INTRODUCTION 10
1.1 Scope & Objectives 10
1.2 Methodology 12
1.3 Structure 13
2 DATA SUMMARY 14
2.1 Existing data reused 14
2.2 Types and formats 14
2.3 Purpose of the data generation 16
2.4 Expected size of the data 20
2.5 Data origin and provenance 20
2.6 Target Groups 21
3 FAIR DATA 23
3.1 Making Data Findable 23
3.2 Making data Accessible 23
3.3 Making data interoperable 25
3.4 Increase data re-use (through clarifying licenses) 26
4 DATA SECURITY 28
5 ETHICS 29
5.1 Ethics under Horizon Europe Programme 29
5.2 Ethical Considerations In Cop-Pilot Project 29
5.2.1 Involvement of Human Participants 30
5.2.2 Data Management Activities 30
5.2.2.1 Processing of Personal Data 30
5.2.2.2 Processing of non-Personal Data 31
5.2.3 Use of AI Systems 31
6 OTHER ISSUES 33
CONCLUSIONS 34
REFERENCES 35

EXECUTIVE SUMMARY

This Executive Summary presents the core motivations, methodology, key findings, conclusions, and recommendations of the COP-PILOT Data Management Plan (DMP), which articulates a robust framework for handling the diverse data arising from IoT–edge–cloud deployments.

COP-PILOT addresses the pressing challenge of managing high-velocity, heterogeneous data streams generated across multiple sectors and technical layers. Rapid data growth, the already applicable and forthcoming EU regulations (notably the GDPR and the Data Act respectively), and the imperative for interoperability and reuse drive the need for a coherent, enterprise-grade DMP. To tackle this, the project conducted a comprehensive analysis of existing and anticipated data assets—spanning telemetry, sensor logs, configuration files, and derived analytics—across representative use cases. This analysis informed the design of policies and technical controls that align with FAIR (Findable, Accessible, Interoperable, Reusable) principles while embedding privacy-by-design and security-by-default practices.

We developed a standardized metadata schema based on community standards and assigned persistent identifiers to all datasets to guarantee long-term discoverability. Secure storage architecture combines tiered encryption, role-based access controls, and immutable audit trails to protect sensitive telemetry. Provenance capture, implemented via the W3C PROV-O model, ensures full traceability from raw acquisition to processed outputs.

In conclusion, the COP-PILOT DMP establishes a scalable, GDPR-compliant, and FAIR-aligned ecosystem for data stewardship. By integrating standardized metadata, persistent identifiers, robust encryption, and ethical governance, it enables seamless collaboration among stakeholders and paves the way for future extensions. We recommend early deployment of the metadata registry and secure repository prototype, periodic audits of compliance measures, and continual updates to governance policies as new legal or technical requirements emerge. This proactive approach will maximize the impact and longevity of COP-PILOT's data assets while upholding the highest standards of privacy, security, and reusability.

Co-funded by the European Union

LIST OF FIGURES

FIGURE 1 OPEN ACCESS TO SCIENTIFIC PUBLICATION AND RESEARCH DATA IN THE WIDER CONTEXT OF DISSEMINATION AND EXPLOITATION.....25

LIST OF TABLES

TABLE 1: COP-PILOT DATA TYPES.....16

TABLE 2: DATA SETS AND PURPOSES AT PILOTING ACTIVITY LEVEL 17

TABLE 3: COP-PILOT'S TARGET GROUPS AND TYPES OF DATA EACH GROUP CREATES, CONSUMES OR DEPENDS ON.....22

ABBREVIATIONS

IP	Internet Protocol
TCP	Transmission Control Protocol
GA	Grant Agreement
GDPR	General Data Protection Regulation
DMP	Data Management Plan
EU	European Union
EC	European Commission
CPU	Central Processing Unit
5G/SDN	5th Generation / Software-Defined Networking
IoT	Internet of Things
SLA	Service Level Agreement
PU	Public
CO	Confidential
WP	Work Package
API	Application Programming Interface
AI	Artificial Intelligence
LLM	Large Language Model
ML	Machine Learning
UC	Use Case
EV	Electric Vehicle
DER	Distributed Energy Resources
RES	Renewable Energy Sources
UAV	Unmanned Aerial Vehicle
OEE	Overall Equipment Effectiveness
TMF	TeleManagement Forum
ETSI	European Telecommunications Standards Institute
KPI	Key Performance Indicator
TG	Target Group
FAIR	Findable, Accessible, Interoperable, Reusable
CIM	Cloud Infrastructure Management
OSL	OpenSlice
SIF	Secure Integration Fabric
RBAC	Role-Based Access Control

Co-funded by the European Union

1 INTRODUCTION

In today's rapidly evolving technological landscape, the convergence of Internet-of-Things (IoT), edge computing, and cloud platforms is generating unprecedented volumes of diverse data. These data—ranging from raw sensor readings and device telemetry to processed analytics—are central to delivering intelligent, real-time services across smart cities, industrial automation, and healthcare monitoring. Yet, without a coherent framework for their management, these assets risk becoming fragmented, inaccessible, or non-compliant with emerging legal and ethical standards.

The COP-PILOT Data Management Plan (DMP) establishes a comprehensive approach to ensure that all project data are collected, stored, processed, and shared in a manner that maximizes value while mitigating risks. It outlines guidelines and tools to make data Findable, Accessible, Interoperable, and Reusable (FAIR). It, also, considers, to the extent relevant, privacy-by-design and data minimization, hence, aligning with the requirements of the General Data Protection Regulation (GDPR), while promoting access and use of data in line with the goals of the Data Act . By prescribing metadata standards, persistent identifiers, provenance tracking, secure storage, and governance protocols, the DMP transforms raw data streams into well-governed, high-quality assets suitable for cross-sector collaboration.

The primary objectives of this plan are to:

- Define the types, formats, and volumes of both existing and newly generated datasets, including their lifecycle and provenance.
- Specify technical measures and policies for secure storage, controlled access, and long-term preservation.
- Detail metadata schemas and identifier schemes to support discovery and interoperability.
- Describe ethical review processes and governance structures for data sharing, especially in AI-driven or high-risk contexts.

By articulating these elements in a single, self-contained document, the COP-PILOT DMP serves as a roadmap for project partners and stakeholders to manage data consistently and effectively throughout the project's duration and beyond. It also lays the groundwork for future extensions, ensuring that as technologies and regulations evolve, the data management ecosystem remains robust, compliant, and aligned with best practices.

1.1 SCOPE & OBJECTIVES

The COP-PILOT Data Management Plan (DMP) encompasses all data-related activities across the project's IoT–edge–cloud use cases, from initial acquisition through long-term preservation and reuse. It applies to datasets generated, processed, and consumed by consortium partners, including both existing legacy archives and novel data streams produced during pilot deployments. The plan addresses technical, organizational, legal, and ethical dimensions, ensuring that every phase of the data lifecycle adheres to best practices and regulatory requirements.

Scope

- **Data Lifecycle Coverage:** All phases (collection, pre-processing, analysis, storage, sharing, archiving, and destruction) of structured, semi-structured, and unstructured data.

- **Data Types**:
  - Raw sensor and telemetry logs from IoT and edge devices
  - Configuration files and device metadata
  - Intermediate and final analytics outputs (e.g., AI model artefacts)
  - Personal and quasi-identifiable data collected under GDPR constraints
- **Technological Layers**: Sensors and actuators at the IoT layer; real-time processing and aggregation at the edge; scalable storage and advanced analytics in the cloud.
- **Stakeholders and Roles**: All project partners, including data producers, processors, repository administrators, and end-user researchers; governed by defined roles and responsibilities.
- **Regulatory and Ethical Boundaries**: Compliance with GDPR, local data-protection laws, and ethical guidelines for high-risk AI; procedures for secure handling of sensitive datasets.

Objectives

1. **FAIR Compliance**
   - Assign persistent identifiers and apply standardized metadata schemas to make data discoverable and interoperable.
   - Register datasets in a centralized metadata registry for unified search and access.
2. **Security & Privacy**
   - Implement privacy-by-design and security-by-default controls, including role-based encryption, anonymization, and audit trails.
   - Define access levels and consent management workflows for personal and quasi-identifiable data.
3. **Provenance & Quality**
   - Capture end-to-end provenance using W3C PROV-O to ensure traceability of data transformations.
   - Enforce validation checks and quality metrics at each processing stage.
4. **Governance & Sharing**
   - Establish Data Access Committees and ethical review procedures to vet sharing requests, especially for AI and sensitive data.
   - Define licensing models and access conditions to balance openness with protection.
5. **Sustainability & Preservation**
   - Specify archival formats, retention schedules, and repository technologies to guarantee long-term usability.

- Plan for regular audits and updates to the DMP as legal, technical, and organizational contexts evolve.

By clearly delineating its scope and objectives, the COP-PILOT DMP ensures that all partners share a unified vision for managing data responsibly, securely, and effectively, maximizing its value both during the project and beyond.

1.2 METHODOLOGY

To ensure a robust, compliant, and practical Data Management Plan, COP-PILOT adopted an iterative, stakeholder-driven development process that aligns with EU guidance on FAIR data, GDPR, and ethical AI. The following steps were performed:

1. Requirements Gathering & Policy Alignment

- Reviewed the Grant Agreement, Horizon Europe Data Management Guidelines, and the AI Act to extract DMP obligations and identify relevant legal, ethical, and technical standards.
- Surveyed existing project documentation (Part B “Implementation methodology” and initial ecosystem requirements) to align data-management goals with platform use-case scenarios.

2. Data Inventory & Classification

- Distributed a structured questionnaire to all piloting clusters and third-party partners, eliciting details on existing and planned datasets (types, formats, volumes, intended use).
- Consolidated responses into a centralized catalogue and classified every dataset as personal, quasi-identifiable, or non-personal to determine appropriate handling protocols.

3. Gap Analysis & Standards Selection

- Compared the current data-handling practices against FAIR principles and GDPR requirements to identify gaps in metadata, security controls, and governance.
- Selected community-endorsed schemas (e.g. Dublin Core, DataCite) and the W3C PROV-O model for metadata and provenance tracking.

4. Technical & Organizational Protocol Definition

- Defined technical measures—end-to-end encryption (TLS 1.2+, AES-256 at rest), Keycloak-based RBAC, immutable audit trails—to secure data in transit and at rest.
- Outlined organizational workflows for data access, sharing, and deletion, appointing PEC-authorized administrators and Data Access Committees.

5. Ethical & Legal Assessment

- In close collaboration with Task 1.3 Legal, Ethics, Gender Balance and SSH Monitoring, a primary ethics review was conducted—including bias, safety, and consent considerations—for all AI-driven use cases, invoking GDPR Article 89 exemptions where appropriate.

- An independent ethics advisor has been identified and will be appointed to support the consortium; (the contract is under preparation at the time of the submission of this deliverable). The role of the ethics advisor is to provide an independent evaluation based on key deliverables that report validation results and outcomes from piloting UCs, and in terms of compliance to EU GDPR rules.

6. Review & Validation

- Held multi-partner workshops to validate the DMP's feasibility with WP2's metadata registry and WP3's secure integration fabric prototypes.
- Incorporated feedback from the consortium's legal, ethics, and technical experts through two peer-review cycles.

7. Iterative Updates

- Committed to a bi-annual revision cycle: each update will enrich the dataset catalogue, refine metadata and security protocols, and document new consent or governance measures.

By combining comprehensive stakeholder consultation, alignment with EU standards, and rigorous technical and ethical assessments, this methodology ensures that COP-PILOT's DMP remains actionable, compliant, and fully integrated with the project's evolving pilots and platform architecture.

1.3 STRUCTURE

The current section (**Section 1 – Introduction**) provides the scope, objectives, and methodology for the Data Management Plan (DMP), establishing a comprehensive approach for handling all project data.

Section 2 - Data summary, provides an overview of the datasets the project will reuse and generate, detailing their types, formats, purpose, expected size, origin, and target groups.

Section 3 - FAIR data, describes the strategies and measures to ensure project data is Findable, Accessible, Interoperable, and Reusable (FAIR) by adhering to open standards and clear licensing.

Section 4 - Data Security, this section outlines the comprehensive technical and organizational safeguards implemented to protect all project data, including encryption, access control, and backup procedures.

Section 5 - Ethics, details the project's adherence to ethical principles and legal regulations under the Horizon Europe Programme, with specific considerations for data management, human participants, and the use of AI systems.

Section 6 - Other issues, addresses the need for partners to comply with national and funder-specific data management policies, particularly for those based in the UK and Switzerland.

In the **conclusions section** the DMP's accomplishments are summarized, and the role of this deliverable as a living document to be updated every six months is confirmed. Also the plan for future revisions to maintain compliance and relevance is outlined.

Co-funded by the European Union

2 DATA SUMMARY

This section provides an overview of the datasets and other digital artefacts that COP-PILOT will (re)use, generate, and share during the project's lifetime.

2.1 EXISTING DATA REUSED

Data will be collected continuously throughout all COP-PILOT activities. In addition to live telemetry from the four large-scale piloting clusters (energy, smart buildings, agriculture, manufacturing), retrospective datasets—both legacy pilot data and benchmarks supplied via open calls—will be curated and reused. For every dataset contributed by third-party participants, acquisition, anonymization, and licensing protocols will be performed ensuring full GDPR and grant-agreement compliance. Piloting partners have committed to updating and enhancing their legacy datasets, and any data provided through open calls must satisfy standardized metadata and quality requirements. To safeguard data sovereignty, sharing will be limited to EU consortium members, non-EU consortium members from United Kingdom and Switzerland which provide an adequate level of data protection as recognized by the issued Data Protection Adequacy Decisions [1] and formally contracted third-country partners (if any, from the open calls); any further non-EU transfers will require explicit contractual authorization and additional security measures.

COP-PILOT will also build upon and extend several existing data sources and repositories:

- **Open IoT/edge benchmarks and datasets**
  - Publicly available IoT sensor streams (e.g. smart-building, precision-agriculture) for platform integration tests.
  - Standard edge-computing performance traces (e.g. CPU/memory metrics) from e.g., P2CODE and ACROSS HEU projects.
- **Telecom and network telemetry**
  - Historical 5G/SDN network logs and SLA records contributed by project partners (INTRA, OTE, TATA).
- **SLA templates and violation datasets**
  - Pre-trained forecasting models and synthetic SLA-violation data generated in AI REGIO and ACROSS HEU projects.
- **Third-party pilot data**
  - Select datasets from previous H2020/HEU pilots (e.g., FIDAL, P2CODE) made available under open licences for cross-validation.

2.2 TYPES AND FORMATS

ARTEFACT TYPE	EXPLANATION	WP#	FORMAT
Research Items	**Deliverables**: The project will produce several deliverables uploaded on the project website.	WP1-WP7	MS Word (.doc/.docx), Adobe PDF (.pdf), MS Excel (.xls, .xlsx), Comma - separated values (.csv), Hypertext Mark -up Language(.html), JPEG

	These deliverables are either public (PU) or confidential (CO). **Scientific publications**: COP-PILOT partners will produce scientific publications that will be made publicly available for the wider audience. **Other dissemination and communication publications**: In the scope of WP7, other publications will be produced like website pages, promotional materials, press releases, website news, and blogs		(.jpeg, .jpg), GIF (.gif), PNG (.png), MPEG-4 (.mp4)
Software	Code, APIs, microservices, libraries, dashboard	WP3-WP4	More common formats for software-related data include code files (e.g., Java, Python, C++, etc.), configuration files (e.g., YAML, JSON, XML, etc.), database files (e.g., SQL, NoSQL, etc.), log files (e.g., text, CSV, JSON, etc.), and various types of binary files (e.g., executables, libraries, etc.), Docker images. The specific format may also depend on the tool and its purpose, as well as any relevant standards or conventions that apply to the particular domain or industry.
Sensor & telemetry streams	Raw and pre-processed time-series from IoT devices (temperature, vibration, location, etc.)	WP4, WP5	SON (MQTT over TCP), CSV, Parquet
Context data	Semantic representations of entities and relationships managed by FIWARE Context Brokers	WP2, WP4	NGSI-LD (JSON-LD)
Orchestrator logs & metrics	Audit trails, API call logs, performance counters from ServOrch & InfraOrch	WP3, WP5	JSON, YAML, plain text
Service blueprints & policies	High-level service specifications and TMF-based policy definitions	WP2, WP4	YAML, XML
Machine-learning artefacts	Trained forecasting models for SLA prediction	WP3, WP5	ONNX, Pickle (.pkl), HDF5

Pilot evaluation datasets	Measurements from large-scale use cases (latency, throughput, energy consumption, etc.)	WP5	CSV, JSON
Personal Data	Personally identifiable information collected for project administration, partner and pilot-site contacts, open-call applicants, consent records, user-survey responses and operator logs. This includes names, email addresses, organizational roles, demographic attributes, and any free-text feedback where individuals may self-identify.	WP1, WP6, WP7	CSV, JSON, XLSX, DOCX (stored encrypted/pseudonymized)

Table 1: COP-PILOT Data types

2.3 PURPOSE OF THE DATA GENERATION

COP-PILOT's data collection strategy is designed to fuel every phase of the project's objectives (from platform innovation to large-scale validation and market exploitation) by capturing rich, multi-level telemetry and contextual information across the IoT–edge–cloud continuum. The primary purposes of data generation are as follows:

- **End-to-end platform development and validation**: Detailed telemetry from ServOrch and InfraOrch components (API calls, resource metrics, orchestration logs) and semantic context updates from FIWARE Context Brokers will enable continuous refinement of the hierarchical orchestration architecture. By analyzing real-time performance data, bottlenecks can be identified and addressed, ensuring the platform meets the KPI targets for scalability (e.g., $\mathcal{O}(log\ N)$ orchestration complexity) and responsiveness (e.g., < 1 ms at extreme edge).

1. **Intelligent automation and SLA preservation**: Data from smart SLA-forecasting models (historical violation logs, prediction outputs, and reconfiguration actions) will feed closed-loop AI workflows that proactively detect and remediate SLA breaches. Generating and analysing these datasets will validate the accuracy (MAPE < 10%, precision/recall > 80%) and latency (anticipatory 2–5 s) of predictive algorithms, driving continuous improvement of zero-touch automation.
2. **LLM-driven user interaction**: Multimodal data (free-text service specifications, uploaded artefacts, and user dialogues) will be logged to train and fine-tune the LLM-UI plugin. This real-world usage data will help reduce onboarding times from tens of minutes to under one minute, ensuring the user interface evolves in step with developer and end-user needs.
3. **Pilot cluster performance and impact assessment**: Large-scale measurements from the four piloting clusters (CL1-CL4) – including latency, throughput, resource utilization, and energy consumption – will be collected to validate cross-sector applications. These datasets will underpin quantitative evaluations of market potential, environmental footprint, and societal benefits, informing both technical refinements and exploitation strategies. Table 3 below indicates a first iteration on data sets and purposes at piloting activity level. While the goal of the table if to provide an up-to-date overview of the data processing activities occurring during the project at the piloting activity level additional legal and ethical considerations concerning the CoP

Table 2: Data Sets and Purposes at Piloting Activity Level

PILOTING CLUSTER	USE CASES	DATA	PURPOSE
Cluster 1: Business integration in mining	UC 1.1: IoT Mining Seismics	Rocksigma private data	Test E2CCC performance facilitating new business contracts.
	UC 1.2: Logistics IoT	Thingwave private data	Test E2CCC performance facilitating new business contracts.
	UC 1.3: Condition Monitoring and Predictive Maintenance in Mining	Predge and Hosch private data	Test E2CCC performance facilitating new business contracts.
	UC 1.4: IoT-Edge-Cloud-Continuum for Digital Mines	Dummy data to validate evaluation of the E2CCC. based on Arrowhead and ColonyOS integration	Validation of E2CCC for digital mines suitability in UC1.1-1.3
Cluster 2: Sustainable mobility for smart city scenarios including a sustainable campus, industrial park, and smart port	UC 2.1A: UC 5G-Connected Radars for Traffic Classification and Vehicle Counting	Radar localization and orientation, vehicle counts for the targeted street of each radar, including timestamped classifications, velocities and lanes.	The data collection aims to enable smarter urban mobility management, improved traffic flow, enhanced road safety, and more sustainable city planning.
	UC 2.1B: Flood warning and mitigation system through radar sensing	Record of timestamped water levels at the measured locations.	The collected data will enable early warnings, faster response, and more effective flood mitigation to protect people, infrastructure, and the environment.
	UC 2.2: Smart Resources Management in the UPV campus	Timestamped waste levels at the measured locations.	The purpose of data collection is to enable accurate tracking, optimized waste collections, and improved sustainability through smarter resource management.
	UC 2.3: Maritime traffic monitoring and berthing assistance	Vessel position and orientation, relative position between vessel and dock and crane position and status.	The collected data will enhance improve berthing safety, prevent collisions and optimize overall port operation.

Cluster 3E: Edge intelligence for increasing grid reliability in RES-rich, cross-sector coupled distribution grids	UC 3E.1: Harvesting in real-time flexibility from active electricity distribution grids	Real-time telemetry data from DERS, including simulated and emulated loads, generation, and storage units. This includes power flows, voltage levels, der status, and environmental conditions.	The collected data will be used to dynamically estimate the flexibility potential of the distribution grid, optimize asset control, simulate active grid conditions, and support SLA-driven service orchestration.
	UC 3E.2: Ensuring Uninterruptible Power Supply for Fast EV Chargers	Real-time voltage and current measurements from the charging stations. Historical data from charging sessions (time of arrival, charging duration, energy consumption, and logs of events).	The data will be used to develop edge AI models to predict the energy demand at each charging station and predictive maintenance per charging station.
	UC 3E.3: Predictive Maintenance and Monitoring of Anaerobic Digestion in a Biogas Plant	Operational and environmental datasets will be collected, including: (i) gas composition data: ch_4, co_2, h_2s, o_2, h_2 (via awiflex gas analyzer), (ii) ph levels: measured using cps11e ph sensor, (iii) temperature and feedstock data, (iv) electrical output performance, (v) maintenance logs and event histories, and (vi) Synthetic data generated from digital twin simulations for performance forecasting and model validation.	Purpose of data collection is to (i) monitor and optimize the anaerobic digestion process in real-time, (ii) enable predictive maintenance to prevent failures, (iii) forecast biogas-based electricity generation, and (iv) improve grid integration of renewable energy these activities are central to the use case's objective of increasing system reliability, energy efficiency, and operational resilience.
Cluster 3A: AgriTech Transformation and Sustainability Initiative	UC 3A.1: Integrated Precision Agriculture and Crop Monitoring	Environmental data from weather stations (temperature, humidity, solar radiation, etc.) UAV multispectral imagery (NDVI, and vegetation indices, crop maps). Satellite data from Copernicus services. Plant wearable sensor data (anti-nutrient levels, stress indicators). Operational logs and traceability information (spraying actions, input applications, etc.).	The purpose of data collection is to enable real-time crop monitoring and targeted intervention, supporting increased sustainability and operational efficiency.

	UC 3A.2: Advanced AgriRobotics for Autonomous Intervention	Camera data (RGB images of plants and weeds), GPS/positioning data for precise navigation in fields, environmental data (e.g., moisture, temperature, wind) to decide where and when to act, crop growth data to identify the right timing for interventions.	To allow the robotic platform to move autonomously, perform targeted spraying and act accurately and efficiently without human help.
	UC 3A.3: Secure Data Management and Interoperability	Batch identifiers, timestamps, GPS location data, operator ID hashes, packaging metadata, and transport event telemetry.	To ensure transparent, secure and interoperable tracking of agri-products from field to shelf, improving traceability, compliance and trust across the supply chain.
	UC 3A.4: Smart Logistics and Supply Chain Optimization	GNSS coordinates, temperature and humidity logs, vehicle identifiers, delivery status events, and routing metadata.	To improve delivery efficiency, ensure product quality during transport and provide real-time visibility into the agri-food supply chain.
Cluster 4: Integrated IoT Solutions for Enhancing Sustainability & Efficiency in Agriculture, Recycling, Manufacturing	UC 4.1: Recycling, Maintenance, and Logistics of IoT sensors	Kit data for IoT Sensors. Location Data for IoT Sensors Kits. Logistics Data for IoT Sensors Kits. Status Data for IoT Sensors Kits.	Tracking the details, location, logistics, and operational status of reusable IoT sensor kits used in patient care, the dataset supports efficient management within the healthcare supply chain, extending device lifecycles, timely maintenance, and sustainable reuse.
	UC 4.2: Water Utilisation Efficiency	Satellite imagery (e.g. Optical, SWIR, TIRS). Historical weather data. Predictive weather data. Topographical information. Historical soil moisture content trends from satellite data. Historical Water Stress Index data.	To support the sensorless estimation of high-resolution soil moisture by integrating satellite imagery, weather data, and IoT sensor metadata. It enables machine learning models to predict topsoil and subsurface moisture content at 30-meter resolution, enhancing real-time monitoring and sustainable agricultural practices.

	UC 4.3: Sustainable optimized Winery Production Lines	Production data from IoT sensors, such as machine status, temperature, humidity, bottle count, and line speed. Operational efficiency data, covering Overall Equipment Effectiveness (OEE) metrics, downtime logs, error reports, and maintenance history. User interaction data, such as manual entries from operators regarding maintenance reports, production adjustments, and quality control inputs.	The primary goal of data collection of this Use Case is to enhance production monitoring, enable predictive maintenance, and support data-driven decision-making in wineries. Anonymized and aggregated datasets may be used for performance analysis, AI model training, and industry research.

4. **Market analysis and exploitation planning**: Usage metrics, business-case KPIs, and sustainability indicators derived from pilot deployments will inform market studies and business-model canvases. By generating data on service uptake, cost savings, and environmental impact, COP-PILOT will craft robust exploitation roadmaps and engage stakeholders with concrete evidence of value.
5. **Dissemination, standardization, and community building**: Curated datasets, software artefacts, and performance reports will be published under FAIR-compliant licences to demonstrate the platform's capabilities. Open sharing of benchmark data (e.g., SLA-violation logs, orchestration traces) and TMF/ETSI-aligned models will catalyse community adoption, foster contributions to standards bodies (ETSI OSL/TFS/ZSM, TMF), and attract third-party piloting use cases through transparent, data-driven evidence of COP-PILOT's value.

2.4 EXPECTED SIZE OF THE DATA

By project end (Month 36), COP-PILOT anticipates:

- **Raw and aggregated IoT/edge telemetry**: ~600 GB
- **Orchestrator logs & context snapshots**: ~200 GB
- **ML model artefacts and intermediate datasets**: ~50 GB
- **Pilot evaluation results**: ~100 GB
- **Dissemination materials and code repositories**: ~50 GB

Total (approx.): **1 TB** of structured and unstructured data.

2.5 DATA ORIGIN AND PROVENANCE

Accurate recording of data origin and provenance is vital for COP-PILOT to guarantee transparency, reproducibility, and compliance with ethical, legal, and contractual obligations. Provenance metadata also ensures proper attribution, clarifies data ownership, and underpins FAIR-compliant sharing and reuse. COP-PILOT classifies data sources into the following categories, each with its own provenance requirements:

1. **In-project generated data**

- **Description:** Raw and processed data produced directly by COP-PILOT components—e.g. telemetry from ServOrch and InfraOrch, LLM-UI interaction logs, and smart-SLA forecasting outputs.
- **Provenance capture:** Each record is tagged with a unique dataset identifier, timestamp, generating component name/version, and responsible partner. Provenance is represented using the W3C PROV-O ontology, linking entities (data artefacts), activities (processing steps), and agents (software modules or personnel).

2. **Partner testbed data**
   - **Description:** Live and historical streams from consortium pilot clusters (energy, smart buildings, agriculture, manufacturing) provided by infrastructure owners (INTRA, OTE, TATA, etc.).
   - **Provenance capture:** Metadata includes the originating testbed site, device or node identifiers, data collection protocol reference, and partner organization. Access restrictions—if any—are documented alongside provenance in the landing page.
3. **Third-party and collaborative pilot data**
   - **Description:** Datasets contributed under open calls by external SMEs, research bodies, or 3rd-party consortia for extended use-case validation.
   - **Provenance capture:** Submission metadata must include contributor details, terms of use, original licence, and any pre-existing provenance records. COP-PILOT augments these with internal processing logs and assigns PIDs for traceability.
4. **Reused open and legacy datasets**
   - **Description:** Publicly available benchmarks, standard datasets (e.g. FIWARE context snapshots, ETSI test vectors), and legacy pilot data from prior EU projects (P2CODE, ACROSS).
   - **Provenance capture:** Original source citations (DOI or URL), licence statements, and date/version of the source snapshot are recorded. Transformation steps (e.g. format conversions, filtering) are logged as provenance activities.
5. **Future primary and externally collected data**
   - **Description:** Data acquired through new partnerships, user surveys, or experiments with non-consortium organizations (e.g. power utilities, agricultural cooperatives).
   - **Provenance capture:** A data-collection will be conducted according to legal and ethical polices.

By embedding these provenance mechanisms, COP-PILOT ensures that all data remains verifiable, properly attributed, and ethically managed throughout and beyond the project's lifetime.

Provenance for each dataset will be tracked via metadata (see Section 3), including contributor, date/time, version, and applicable licence.

2.6 TARGET GROUPS

Identifying and segmenting the key target groups for COP-PILOT ensures that data management practices are aligned with the needs, expectations, and constraints of everyone who creates, uses, or governs the project's outputs. By profiling each stakeholder group (understanding their preferred formats, access requirements, and potential concerns) COP-PILOT can design workflows, repositories, and metadata conventions that maximize usability and adoption while minimizing legal, ethical, or technical barriers. This target group-driven approach fosters transparency and trust, promotes cross-domain collaboration, and anticipates challenges in data sharing or reuse before they arise. Ultimately, tailoring the Data Management Plan to these stakeholder profiles will amplify COP-PILOT's impact, sustainability, and uptake across research, industry, and policy communities.

Below is a concise extraction of the COP-PILOT target groups (TG-A through TG-G) and the types of data each group creates, consumes or depends on:

Table 3: COP-PILOT's Target Groups and types of data each group creates, consumes or depends on

Target Group	Data of Interest / Produced
TG-A: Cloud, edge & IoT service providers and users	• Platform telemetry (CPU/memory, network, QoS) from cloud/edge nodes • IoT device streams (sensor readings, event logs) • Service usage metrics (API calls, onboarding times)
TG-B: Critical infrastructure operators and owners	• Operational data from essential services (power, water, transport) including real-time status and incident logs • Policy definitions governing access to facility data • Reliability and security audit trails
TG-C: Public sector (municipalities, regional & national governments, policy makers)	• Aggregated service dashboards for smart-city applications (traffic, public safety, energy usage) • Planning and investment data linked with COP-PILOT's cross-sector pilots • Regulatory compliance reports and policy outcome metrics
TG-D: Knowledge providers (academic/research communities, SDOs such as ETSI, TMF, ISO/IEC, IEEE, etc.)	• Benchmark datasets for AI and IoT experiments (SLA-violation logs, performance traces) • Standards drafts and metadata models (NGSI-LD contexts, TMF data schemas) • Validation studies and published measurement data
TG-E: Technology providers and entrepreneurial ecosystems (SMEs, start-ups, mid-caps)	• Prototype integration artifacts (connectors, SDKs, microservice templates) • Market analysis datasets (usage patterns, business-case KPIs) • Sustainability and impact assessments derived from pilot deployments
TG-F: Civil sector (general public, NGOs, global bodies e.g. Internet Society, ICANN, OECD, UNESCO)	• High-level summaries of pilot results (e.g. environmental and social impact indicators) • Open-access educational materials (infographics, how-to guides) • Anonymized usage statistics to illustrate societal benefits
TG-G: Regulatory bodies (European Commission, EDPS, NRAs, EDPB)	• Compliance datasets demonstrating GDPR and network-security adherence • Audit logs of data handling and consent management • Standardization contributions tracking (e.g. open APIs, data models aligned to ETSI/TMF)

Co-funded by the European Union

3 FAIR DATA

3.1 MAKING DATA FINDABLE

COP-PILOT will adhere rigorously to the principle “as open as possible, as closed as necessary,” not only publishing FAIR-compliant data wherever feasible but also actively enabling the wider community to adopt transparent, reproducible data-sharing practices. To ensure all COP-PILOT outputs are discoverable, interoperable, and reusable, the following measures will be applied:

- **Persistent identifiers and landing pages:** Every dataset, software release or service blueprint will receive a globally unique, persistent identifier (e.g. DOI via DataCite or Handle) and a corresponding landing page in the COP-PILOT catalogue (e.g. Zenodo, GitHub/GitLab or the project portal) where metadata, access conditions and version history are clearly documented.
- **Standardized metadata schemas:** Metadata records will conform to both Dublin Core and DataCite schemas (including title, creator/orcid, version, licence, keywords, abstract, funding reference, and provenance), ensuring machine-actionable discovery and interoperability across repositories.
- **Rich keyword tagging:** Domain-specific controlled vocabularies (IoT, edge computing, SLA forecasting, etc.) and EDAM/SMART ontologies will be used to tag each dataset, maximizing findability and facilitating semantic search and cross-domain reuse.
- **Transparent versioning:** Every change—whether a minor update or major release—will be captured as a new version with its own PID and changelog, allowing users to trace the evolution of data products over the project lifetime.

These provisions guarantee that COP-PILOT’s data assets remain accessible, well-documented, and primed for reuse by researchers, industry stakeholders, and standardization bodies long after project completion.

3.2 MAKING DATA ACCESSIBLE

By default, COP-PILOT will publish research outputs “as open as possible, as closed as necessary,” balancing transparency with ethical, legal, and security obligations. Accessibility measures will be applied at three complementary levels:

Repository level

- **Trusted repository deposits:** All FAIR outputs—datasets, software releases, service blueprints—will be deposited in long-term, discipline-agnostic repositories providing persistent identifiers and versioning. Preferred platforms include:
  - **Zenodo:** General-purpose open-access repository with DOI minting, versioning, and GitHub integration.
  - **GitHub / GitLab:** Hosting for code, container images, and small-scale datasets; releases tagged with DOI via Zenodo integration or Git LFS.
- **Supplementary portals:** When appropriate, aggregated metadata and landing pages will also be published on the COP-PILOT project website or partner data portals, ensuring multiple, sustainable access points.
- **Sustainability considerations:** Repository choices will be re-evaluated annually to confirm continued availability, funding, and compliance with community best practices.

Co-funded by the European Union

Data level

- **Open vs. restricted data:**
  - **Open datasets** will carry CC0 or CC-BY 4.0 licences and be freely downloadable via HTTPS or API.
  - **Restricted datasets** (e.g. sensitive pilot telemetry, personal data) will remain discoverable—metadata will clearly state access conditions—but require authenticated requests under GDPR-compliant procedures.
- **Access conditions & governance:**
  - Access to restricted data will be managed by the Project Coordinator and Technical Coordinator.
  - Authentication will leverage the COP-PILOT Secure Integration Fabric (SIF) and Keycloak identity management.
- **Clear communication of restrictions:** Metadata for every dataset will include license terms and any special conditions (e.g., embargo periods, non-EU transfer prohibitions), ensuring users understand limitations before download.

Metadata level

All COP-PILOT metadata will be openly published under CC0 and include:

- **Descriptive properties:** Title, abstract, keywords, creator names and ORCIDs, date of creation/modification, spatial/temporal coverage.
- **Technical properties:** Data format(s), file size, repository location and PID, version number.
- **Legal & provenance properties:** Licensing terms, funding reference, data origin (e.g. pilot cluster name, partner), processing steps and quality-control notes.
- **Access properties:** Access level (open/restricted), DAC contact point, applicable embargo or expiry dates.

Metadata records will conform to Dublin Core and DataCite schemas and will be harvested via OAI-PMH or RESTful APIs to facilitate automated discovery and interoperability.

Research outputs & publications

In line with the EC Guidelines on Open Access to Scientific Publications and Research Data in Horizon Europe, COP-PILOT will adopt a **Gold and Green Open Access** strategy:

- **Gold OA:** Encouraged for high-impact journals, ensuring immediate open availability under compliant licences.
- **Green OA:** Self-archiving in institutional or subject repositories when publisher policies allow, subject to any agreed embargo.

A deterministic **Result Classification Procedure** will be used to label every project deliverable—white paper, journal article, anonymous usage dataset—as **Public** or **Non-Public**. Public results will follow the open-access route; non-public results will be restricted in accordance with consortium and funder rules.

The diagram in Figure 1 illustrates how COP-PILOT transforms raw research outputs into either openly shared knowledge or protected intellectual property through two parallel decision processes.

Co-funded by the European Union

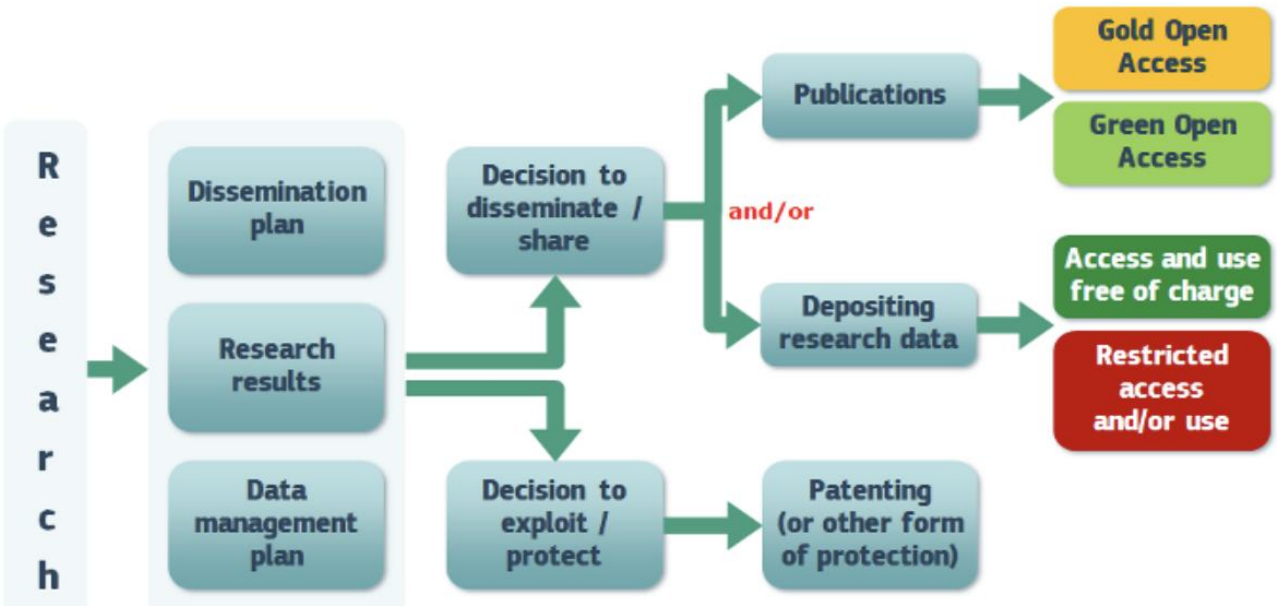

Figure 1 Open access to scientific publication and research data in the wider context of dissemination and exploitation.

At the heart of the workflow are three foundational documents: the Data Management Plan, which specifies how datasets will be collected, described, and preserved; the Dissemination Plan, which lays out the channels and timing for communicating findings; and the Research Results themselves—the experimental data, software, and insights produced during the project. Once these inputs are in place, a formal decision to disseminate or share is made. If the choice is to publish, the consortium selects between Gold Open Access (immediate, publisher-hosted open access under a liberal license) or Green Open Access (self-archiving in a repository, potentially after an embargo). If the output is a dataset, it is deposited in a trusted repository where it may be made freely accessible or, if necessary for privacy, security, or commercial reasons, placed under restricted access with clear metadata describing conditions of use. In parallel, the consortium may decide to exploit or protect certain innovations; in that case, novel methods, algorithms, or devices move toward patenting or other forms of legal protection rather than immediate public release. This dual-path approach ensures that COP-PILOT maximizes both the openness and impact of its research where appropriate, and the strategic safeguarding of its most valuable intellectual assets.

All of these measures ensure that COP-PILOT's research outputs remain discoverable, accessible, and reusable—maximizing impact while upholding ethical, legal, and security requirements.

3.3 MAKING DATA INTEROPERABLE

COP-PILOT places interoperability at the core of its data strategy to accelerate innovation, enable seamless cross-domain integration, and support reuse and validation of research outputs. To achieve a truly connected data ecosystem—spanning IoT devices, edge platforms, network services, and cloud infrastructures—the following measures will be implemented:

Adoption of open standards for data models and APIs

- **NGSI-LD (JSON-LD):** All context data exchanged via FIWARE Context Brokers will conform to the NGSI-LD information model, ensuring a shared semantic foundation across IoT, edge, and cloud domains.
- **TMF OpenAPI interfaces:** Service and resource management APIs (north- and south-bound) will be described using the OpenAPI specification, enabling language-agnostic client generation and automated integration into diverse toolchains.

- **ETSI CIM & ETSI OSL profiles:** Infrastructure and network telemetry will follow ETSI Cloud Infrastructure Management (CIM) and OpenSlice (OSL) data schemas, guaranteeing compatibility with European standardization initiatives.

Use of Linked Data principles

- **RDF/JSON-LD serializations:** Wherever graph-structured relationships are required (e.g., asset registries, service dependency graphs), data will be published in RDF or JSON-LD formats.
- **IRI-based entity linking:** Devices, services, and stakeholders will be identified by HTTP-IRI PIDs, facilitating federated queries and integration with external knowledge graphs (e.g., Wikidata, schema.org).

Controlled vocabularies and ontologies

- **Domain vocabularies:** IoT and edge concepts will leverage existing ontologies (SSN/SOSA for sensors/observations, QoS4IoT for service quality) and TM Forum (TMF) data models for business entities and policies.
- **Schema.org annotations:** Public-facing artefacts (landing pages, documentation) will embed Schema.org metadata to enhance search engine discoverability and interoperability with web-scale data harvesters.
- **Ontology alignment:** Project-specific extensions will be mapped to mainstream EOSC vocabularies or community-endorsed ontologies; equivalence and subclass relationships will be declared via OWL or SKOS mappings.

Open-source toolchains

- All interoperability components (NGSI-LD brokers, API gateways, RDF triplestores) will be deployed using OSS projects—FIWARE, Apache Jena, Swagger UI—ensuring transparency and ease of deployment in partner infrastructures.
- CI/CD pipelines will validate schema compliance and run automated integration tests to detect breaking changes early.

Provision of mappings and transformation scripts

- When proprietary or legacy data formats must be ingested (e.g., CSV logs from third-party pilots), transformation scripts (Python, Java) and mapping documentation will be provided to convert data into the canonical NGSI-LD/RDF models.
- Versioned XSLT or JSON-LD context files will document all transformations, ensuring traceable ETL processes.

Referencing and provenance

- All reused datasets will include explicit provenance metadata (using W3C PROV) pointing to original sources and licences.
- References to external standards (e.g., OpenAPI v3.0, RDF 1.1, Dublin Core) will be embedded in metadata fields to facilitate automated compliance checks.

By rigorously applying these standards and practices, COP-PILOT ensures that its data can flow effortlessly across technical, organizational, and geographic boundaries—laying the groundwork for large-scale, cross-sector innovation.

3.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES)

COP-PILOT is committed to maximizing the value and impact of its data by facilitating broad reuse for validation, extended analysis, and novel application development across the IoT–edge–cloud continuum. The following provisions will ensure that COP-PILOT datasets and artefacts remain transparent, well-documented, and readily reusable by external researchers and industry stakeholders:

Methodological transparency

- Every dataset will be accompanied by a **README** detailing the data-collection protocol (device types, sampling rates, environmental conditions), preprocessing/cleaning steps, and any assumptions or filters applied.
- A **Data Processing Report** will outline transformation pipelines, aggregations, and statistical methods, enabling others to reproduce derived metrics.

Code and workflow documentation

- All analysis scripts, ETL pipelines, and model-training code will be versioned in Git repositories (GitHub/GitLab), with inline comments, dependency manifests (e.g., requirements.txt, Dockerfiles), and usage examples.
- **Workflow definitions** (e.g., CWL, Nextflow, or GitHub Actions) will be provided to automate end-to-end data processing, from raw ingestion to analysis outputs.

Licensing for maximal reuse

- **Data** will be licensed under CC0 or CC-BY 4.0, selecting the option that imposes the fewest restrictions while respecting privacy and security constraints.
- **Software** will adopt permissive licences such as Apache 2.0 or MIT to allow integration into both open-source and proprietary toolchains.

Post-project availability

- Upon project completion, all FAIR outputs will remain accessible in their hosting repositories.
- **Restricted datasets** (e.g., sensitive infrastructure telemetry) will be made available to bona fide researchers after DAC approval and under a Data Use Agreement, with clear metadata indicating any embargo or access conditions.

Provenance capture and linkage

- Provenance metadata will follow the W3C PROV-O ontology, capturing key entities (datasets), activities (processing steps), and agents (software or personnel).
- Structured PROV records (RDF/JSON-LD) and human-readable provenance logs will link each data artefact to its source, transformations, and responsible parties.

Quality assurance and validation

- A **Data Quality Plan** will define checks for completeness, consistency, and outlier detection, with automated validation scripts run in CI pipelines.
- **Audit reports** will document any anomalies found, corrections made, and rationale for data exclusions.

By embedding these measures throughout the data lifecycle, COP-PILOT ensures that its datasets and software artefacts can be confidently reused, extended, and integrated—accelerating scientific discovery and industrial innovation long after the project's end.

4 DATA SECURITY

COP-PILOT implements a comprehensive set of technical and organizational safeguards to protect all project data—especially any personal or sensitive information—against unauthorized access, unlawful processing, alteration, loss, or destruction. Key measures include:

Encryption at rest and in transit

- All data stored in the COP-PILOT Secure Integration Fabric (SIF) and partner cloud workspaces are encrypted at rest using AES-256.
- All communications—whether between IoT devices, edge nodes, brokers, or user interfaces—occur exclusively over TLS 1.2+ (HTTPS), ensuring end-to-end confidentiality and integrity.

Credential and access management

- User credentials are never stored in plaintext; passwords and tokens are protected using one-way hashing (e.g. bcrypt) and secure vaulting.
- Role-based access control (RBAC) enforces least-privilege: only designated administrators may create, modify, or delete datasets; read-only or download-only access can be granted to other authenticated users.
- Authentication and authorization leverage Keycloak SSO integrated with the SIF, enabling fine-grained consent management and audit logging.

Backup, versioning, and disaster recovery

- Automated snapshot backups are performed, with full off-site replicas retained for 30 days.
  - Every data modification triggers a version checkpoint, allowing rapid rollback to any prior state.
  - In the event of catastrophic infrastructure failure, the most recent snapshots are restored automatically within minutes, ensuring continuity with minimal data loss.

Controlled long-term preservation

- Datasets approved for open-access publication are migrated to Zenodo (hosted on CERN EOS), which maintains two independent file copies on separate disk arrays and dual MD5 checksums (Invenio and EOS level) for real-time corruption detection and recovery.
- Repository selection is guided by stringent security certifications, redundancy, and community trust.

Audit, monitoring, and compliance

- All data access and administrative actions are logged with immutable audit trails, supporting GDPR-compliant record-keeping and forensic review.
- Vulnerability assessments, penetration tests, and compliance audits may be conducted to ensure ongoing adherence to EU data-protection and funding requirements if needed.

Through these layered security controls—encryption, strict access governance, continuous backups, and trusted archival repositories—COP-PILOT guarantees that data remain both robustly protected and reliably available throughout the project lifecycle and beyond.

5 ETHICS

This chapter briefly sets the scene in terms of applicable regulations on ethics under Horizon Europe Programme and provides an overview of main ethical considerations in COP-PILOT project.

5.1 ETHICS UNDER HORIZON EUROPE PROGRAMME

In all EU-funded activities, ethics play a central role throughout the entire research process from beginning to end. Under Horizon Europe Programme[1] which is the EU's key funding programme for research and innovation it is emphasized that research excellence can only be achieved through ethical research with the application of fundamental ethical principles and legislation. To this end it is made clear in article 19 of Horizon Europe Regulation [2] that all activities carried out under the Horizon Europe Programme must be in compliance with ethical principles that are relevant to Union, national and international law [3]. How to achieve such compliance is further explained in the latest Horizon Europe Programme Guide version 1.5 published by the European Commission in 2022 [4] ("Guide") by clarifying that an Ethics Appraisal Procedure will be implemented inclusive of an ethics review to be conducted before the start of the project as well as ethics checks, reviews and audits throughout the project.

Ethics review include compliance with ethical rules and standards, relevant European legislation, international conventions and declarations, national authorizations and ethics approvals, proportionality of the research methods and the applicants' awareness of the ethical aspects and social impact of their planned research [5]. The Guide clarifies that an ethics assessment involve investigations on ethical principles as well as examinations of legal compliance. Thus, projects like COP-PILOT that help EU with reaching its digital decade targets need to adopt all necessary procedures and measures to comply with the legal requirements arising from the applicable EU legislation including but not limited to GDPR [6], AI Act [7], Data Act [8] and Data Governance Act [9] along with regularly assessing ethical issues posed by the project activities.

In this context an overview of ethical considerations in COP-PILOT project is provided below.

5.2 ETHICAL CONSIDERATIONS IN COP-PILOT PROJECT

COP-PILOT tackles key hurdles that the EU must overcome to meet its digital ambitions—namely, complex interoperability, the exponential growth of IoT deployments, and limited transparency in data governance. By creating a market-ready, cross-sector, multi-domain computing environment via a Collaborative Open Platform piloting framework, COP-PILOT recognizes the ethical implications of advanced IoT, edge, and cloud services.

All consortium partners are committed to upholding European values, rights, and principles throughout every project activity. To this end, Task 1.3—Legal, Ethics, Gender Balance and SSH Monitoring—led by our in-house legal and ethics expert, systematically examines the ethical, legal, and societal dimensions of COP-PILOT. This task considers both the project's technical developments and evolving EU regulatory and policy landscapes. Its findings will inform Deliverable

[1] https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/horizon-europe_en

2.1 “Ecosystem Definition and Requirements” (M10), Deliverable 2.2 “COP-PILOT Architecture and Functionalities” (M18), and this living Data Management Plan (D1.2), which will be updated regularly.

Furthermore, an independent ethics advisor, appointed under Deliverable 8.1 “OEI – Requirement No. 1,” will oversee piloting activities—guiding platform design and execution to ensure adherence to the highest ethical standards. The Ethics Advisor will also review the public deliverables that report the outcomes from the validation activities and suggest changes (if required) in compliance to EU regulations and ethical standards.

Because COP-PILOT’s platform spans diverse industries and fosters cross-sector applications, identifying and addressing shared ethical considerations aims to ensure comprehensive, robust governance building on the unified ethics assessment conducted during the proposal phase. Accordingly, an overview of the ethics issues related to COP-PILOT is provided below in a more focused way following the three previously identified sections in the ethics summary report.

5.2.1 Involvement of Human Participants

At this stage of the project (M5), related partners indicated that none of the cluster-related piloting activities involve human participants. Should future phases include volunteers, it will be ensured that any volunteer participants will be informed about the purpose, nature, objective, possible impacts and procedure of the project activities in clear, plain language prior to their actual participation. It will be made explicit that participation is voluntary, based on their free choice and the participants will have the right to withdraw from the project whenever they want without facing any significant consequences. In addition, it will be made sure that the involvement of human participants does not result in any discriminatory practices or unfair treatment of participants. This means that ethical standards and guidelines are intended to be properly applied, regardless of the country in which the research takes place.

5.2.2 Data Management Activities

The EU Data Strategy envisions the EU as a frontrunner in a data-driven society—while upholding safety, security, and fundamental rights. Data protection and privacy are cornerstones of European research ethics, and all COP-PILOT data-management practices strictly adhere to EU and national regulations. Consortium partners employ GDPR’s data-protection principles and a privacy-by-design methodology, reinforced by end-to-end encryption (TLS 1.2+ in transit, AES-256 at rest), role-based access control via Keycloak, and a secure integration fabric for cross-domain transfers.

We collect only the data necessary for each pilot and grant access solely to authorized partners. Any request to share or delete data must be approved by designated project-appointed administrators and is recorded in immutable audit logs. COP-PILOT does not process high-risk categories such as data on minors. When exchanging data with non-EU countries (e.g., the United Kingdom and Switzerland), we rely on their recognized adequacy under the EU’s Data Protection Adequacy Decisions [10]. Further technical specifications— including retention policies and additional security controls—are detailed in Chapter 3 of this Deliverable.

In the following sections a more specified overview of data processing activities under COP-PILOT is explained for both personal and non-personal data.

5.2.2.1 Processing of Personal Data

At this stage, no personal data processing is planned for COP-PILOT’s pilot activities. All pilot-related datasets are catalogued in Table 2, which will be kept current throughout the project. Should any dataset containing personal information be introduced, the partners responsible will process it in full

accordance with GDPR requirements. Personal data will be pseudonymized or anonymized wherever feasible, and robust technical and organizational measures—such as access controls, encryption, and audit logging—will be applied. When exchanging personal data with non-EU countries (e.g., the United Kingdom and Switzerland), their recognized adequacy under the EU's Data Protection Adequacy Decisions [10] will be relied on. Specific safeguards for any future personal data processing will also be documented in subsequent versions of this Deliverable.

According to the Grant Agreement, all personal data processed for the purposes of COP-PILOT will be conducted in accordance with the GDPR. In this respect, as COP-PILOT's activities constitute bona fide research, partners acting as data controllers may invoke GDPR Article 89's research exemptions. In light of this specific Article, provided that appropriate safeguards (e.g., data minimization, pseudonymization, and enhanced security) are in place—and subject to national law—personal data may be repurposed for research or archival use beyond its original collection intent. Similarly, storage periods may be extended, and certain limitations on rectification or erasure may be justified to preserve data integrity for research objectives.

5.2.2.2 Processing of non-Personal Data

Under the EU Data Strategy access to data and the ability to use it are found essential for innovation, growth and competitiveness. COP-PILOT puts utmost importance to working towards improving data access and use of data to foster growth in EU's digital market without compromising privacy and security. For the exploitation of full potential of non-personal IoT data COP-PILOT benefits from the Data Act as it lays down legal and technical foundations for data sharing and interoperability.[2]

Cross-sector collaboration is fostered in areas like energy, agriculture, and smart cities while respecting European values. In addition, COP-PILOT has open-source commitments meaning wherever possible, data, software and models are released to foster innovation. Where appropriate Apache 2.0 or MIT licences and CC0/CC-BY is adopted to maximize reuse while safeguarding proprietary innovations.

5.2.3 Use of AI Systems

In line with its objectives to develop AI based extensions for a better user interface experience and for utilizing intelligence and automation in use cases, COP-PILOT plans to employ AI tools. Based on the Guidelines on the Definition of AI Systems, Article 3(1) of the AI Act, and Recital 12 of the AI Act for a technology to be considered an AI system it must fulfil the criteria outlined in Article 3(1) of the AI Act, which defines AI systems based on main elements. In line with the above-mentioned articles and guidelines, considering the information available at month 4 of the project AI tools employed in COP-PILOT are assumed to be AI Systems based on the definitions in AI Act as the AI capabilities are *machine based*, have different levels of *autonomy* (ML model employed for UC 4.2 is planned to require human intervention as a standard feature whereas ML model employed for UC 3E.1 is planned to either have zero-touch automation or require validation from a human operator) *with different objectives* while *influencing environments* and *generating different outputs* (AI-driven analytics system in UC 4.3 is planned to generate predictions on equipment failures, and detect anomalies whereas in UC 3A.2 ML systems is planned to generate recommendations for agronomists or farmers) *inferring from the input* (mainly from the IoT devices on edge) they receive.

It should be noted that at this stage in the project the available level of detail on the use of AI in piloting activities and the deployed ethics mechanisms differ for clusters and use cases. For example, a primary risk assessment has been planned to address (i) Algorithmic bias and fairness,

[2] To raise awareness in Consortium an internal webinar on "Data Sharing" has been conducted on March 26 by Arthur's Legal, legal and ethics expert in the Consortium.

(ii) Failure modes of real-time AI in edge environments, and (iii) Impact on operator decision autonomy in UC 3E.1. It is further decided that human-in-the-loop configuration is supported, especially during pilot and evaluation phases and model explainability is ensured via visual output and diagnostic metrics on the operator dashboard. While UC 3E.3 employs a similar primary assessment to ensure compliance with ethical AI principles and safety requirements; in UC 3A.1 and UC 3A.2 bias risk is addressed through inclusive training datasets and monitoring performance across diverse farm conditions; transparency is ensured by keeping all algorithmic outputs explainable through dashboards and traceable logs; and human oversight is fully maintained for autonomous recommendations. In UC 1.2 and UC 1.3 AI implementations are transparent, made after vigorous risk assessments and are always handled by a human operator in the end. Regardless of the currently available information and preferred method all project partners commit to maintaining high ethical and legal standards for AI.

Being an EU research project, AI Act is not applicable for COP-PILOT in line with the Article 2(6) of AI Act. However, the AI system usage envisioned within COP-PILOT beyond the specific scope and duration dictated in the related Grant Agreement, should be assessed in line with the risk-based approach of AI Act. In this respect while, the majority of the AI systems under COP-PILOT does not seem to impact fundamental rights, early information shows that in UC 3E.3 the AI system likely falls under "*high-risk*" AI as per Annex III of the AI Act, since it is used to operate critical infrastructure (i.e., a biogas plant integrated into the electricity grid). Even though it does not directly control physical actuators, its role in process reliability and safety justifies this classification. Also, in UC 3E.1 the AI system likely falls under "*high-risk system*" as per the AI Act, since it affects safety-critical infrastructure. At a post project stage, any high-risk AI system will be subject to additional requirements under Articles 8-17 of AI Act and should be continuously monitored.

6 OTHER ISSUES

COP-PILOT brings together 45 partners from 12 countries—10 of which are EU Member States and 2 non-EU countries (the United Kingdom and Switzerland). Partners from each nation must comply not only with Horizon Europe's DMP requirements but also with national and funder-specific data-management policies:

United Kingdom (non-EU): UK-based partners (e.g. Kingston University, University of Bradford, TATA Communications UK) adhere to **UKRI's Common Principles on Data Policy**, which require:

- A detailed, FAIR-aligned Data Management Plan submitted at project start and regularly updated.
- Deposit of research data in approved repositories (institutional or domain-specific), with open licences (CC0/CC-BY) when possible.
- Conformity with UK GDPR and any discipline-specific council guidelines (e.g. EPSRC, NERC) on metadata standards and preservation periods.

Switzerland (non-EU): Swiss partners (e.g. Digital for Planet-D4P, Terraview) follow the **SNSF Guidelines on Open Research Data**, mandating:

- Early planning for data sharing via trusted repositories with DOI assignment and versioning (e.g. Zenodo, SWITCHdrive).
- Controlled access mechanisms for sensitive data, underpinned by documented Data Use Agreements.
- Regular DMP updates through the mySNF portal and compliance with the Swiss Federal Act on Data Protection (FADP).

EU Member States and sectoral rules: Partners in Cyprus, Greece, Italy, Ireland, Germany, Luxembourg, the Netherlands, Portugal, Spain, and Sweden must implement national interpretations of the GDPR and any sectoral mandates (e.g. critical-infrastructure controls for energy telemetry). In particular:

- Energy and utilities partners may need secure enclaves and additional consent procedures for infrastructure-derived data.
- Agriculture and environment partners may follow national open-data portals and specific geospatial metadata requirements.
- Telecoms and ICT partners must align with ePrivacy and network-security regulations.

All third-country and sector-specific requirements will be considered elaborately; local DMP entries will reference relevant national guidelines, ensure alignment with the project's overarching FAIR, security, and ethical principles, and be reviewed at least once per year to accommodate any regulatory changes.

Co-funded by the European Union

CONCLUSIONS

This first version of the COP-PILOT Data Management Plan (D1.2) establishes a solid foundation for the stewardship of all data generated and processed throughout the project's IoT–edge–cloud use cases. It has:

- **Catalogued and characterised** existing and anticipated datasets—ranging from raw sensor telemetry and configuration artifacts to AI-model outputs—and mapped them to their respective purposes in the four piloting clusters.
- **Defined FAIR-aligned metadata schemas** and persistent identifier schemes to guarantee dataset discoverability and interoperability across domains.
- **Specified GDPR-compliant protocols** for handling personal and quasi-identifiable data, including anonymisation, pseudonymisation, and the invocation of Article 89 research exemptions where appropriate.
- **Described robust security measures**, such as end-to-end encryption (TLS 1.2+ in transit, AES-256 at rest), Keycloak-based role-based access controls, immutable audit trails, and secure cross-domain integration via the Secure Integration Fabric.
- **Embedded ethical governance**, appointing a consortium-wide ethics expert (Task 1.3) and an independent ethics advisor (Deliverable 8.1) to oversee data-related activities and ensure compliance with EU values and evolving AI regulations.

To maintain relevance and rigor as COP-PILOT evolves, D1.2 will be **formally updated every six months**. Each revision will:

1. **Incorporate newly generated datasets** with detailed descriptions of formats, volumes, and provenance.

2. **Refine metadata and identifier assignments**, ensuring seamless integration with the metadata registry launched under Deliverable 2.1 "Ecosystem Definition and Requirements" (M10) and the COP-PILOT Architecture specified in Deliverable 2.2 (M18).

3. **Augment security and privacy controls**, reflecting advances in encryption, consent management, and audit-trail technologies.

4. **Extend ethical and legal safeguards**, documenting any new consent workflows, high-risk AI assessments, or updates to GDPR and AI Act compliance measures.

5. **Aligning with emerging standards** from ETSI, TMF, and other SDOs as piloting clusters validate cross-sector interoperability and platform functionalities.

By iterating on this living DMP in lockstep with COP-PILOT's technical milestones and regulatory developments, the consortium ensures that its data assets remain **secure**, **compliant**, **discoverable**, and **reusable**—maximizing both **project impact** and **long-term value** for the European research and industrial communities.

REFERENCES

[1] Commission Implementing Decision (EU) 2021/1772 of 28 June 2021 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by the United Kingdom (notified under document C(2021)4800) (Text with EEA relevance) Official journal of the European Union. L 360/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021D1772 and 2000/518/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland (notified under document number C(2000) 2304) (Text with EEA relevance.) Official journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000D0518

[2] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013

[3] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013 article 19

[4] European Commission, "Horizon Europe (HORIZON) Programme Guide", (Brussels, 01.02.2022) https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_v1.5_en.pdf

[5] European Commission, "Horizon Europe (HORIZON) Programme Guide", (Brussels, 01.02.2022) https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_v1.5_en.pdf

[6] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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 (General Data Protection Regulation). Official journal of the European Union. L 119/1. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official journal of the European Union. L series. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

[8] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). Official Journal of the European Union. L series. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202302854&qid=1730547999853

[9] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). Official journal of the European Union. L 15/21. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868

[10] Commission Implementing Decision (EU) 2021/1772 of 28 June 2021 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of

personal data by the United Kingdom (notified under document C(2021)4800) (Text with EEA relevance) Official journal of the European Union. L 360/1. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021D1772 and 2000/518/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland (notified under document number C(2000) 2304) (Text with EEA relevance.) Official journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000D0518

Anything that is related but not core to the deliverable can go into appendix.